PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

B2

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, C12N 5/10, A61K 48/00, G01N 33/68, C12Q 1/68, C07K 16/18

A1

(11) International Publication Number: WO 95/18225

(43) International Publication Date: 6 July 1995 (06.07.95)

(21) International Application Number: PCT/GB94/02822

(22) International Filing Date: 23 December 1994 (23.12.94)

(30) Priority Data:

9326470.3	24 December 1993 (24.12.93)	GB
9411900.5	14 June 1994 (14.06.94)	GB

(71) Applicants *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB). LEIDEN UNIVERSITY [NL/NL]; P.O. Box 9500, NL-2300 RA Leiden (NL). UNIVERSITY OF WALES COLLEGE OF MEDICINE [GB/GB]; Heath Park, Cardiff CF4 4XN (GB). ERASMUS UNIVERSITY ROTTERDAM [NL/NL]; Burg Ondlaan 50, Postbox 1738, NL-3000 DR Rotterdam (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HARRIS, Peter, Charles [GB/GB]; 65 Freelands Road, Oxford OX4 4BS (GB). PERAL, Belen [ES/GB]; 77 Lock Crescent, Kidlington, Oxford OX5 1HF (GB). WARD, Christopher, James [GB/GB]; 30 Benson Road, Oxford OX3 7EH (GB). HUGHES, James [GB/GB]; 225 Crowley Road, Oxford OX4 1XD (GB). BREUNING, Martin, Hendrik [NL/NL]; Brigantijnstraat 57, NL-1503 BR Zaandam (NL). PETERS, Dorothea, Johanna, Maria [NL/NL]; Zuster Meijboomstraat 267, NL-2331 PH Leiden (NL). ROELFSEMA, Jeroen, Hendrik [NL/NL]; Vijf Meilaan 2006, NL-2321 RR Leiden (NL). SAMPSON, Julian [GB/GB]; 34 Bridge Street, Cardiff CF5 2EL (GB). HALLEY, Dirkje, Jorijntje, Johanna [NL/NL]; van Aerssenlaan 35d, NL-3039 KD Rotterdam (NL). NELLIST, Mark, David [GB/NL]; Noordmolenstraat 57b, NL-3053 RG Rotterdam (NL). JANSSEN, Lambertus, Antonius, Jacobus [NL/NL]; Schokker 37, NL-2991 DJ Barendrecht (NL). HESSELING, Ajenne, Lique, Wilhelma [NL/NL]; Haya van Someren Downerpad 7, NL-3207 DK Spijkenisse (NL).

(74) Agents: NEWELL, William, Joseph et al.; Wynne-Jones, Laine & James, 22 Rodney Road, Cheltenham, Gloucestershire GL50 1JJ (GB).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: POLYCYSTIC KIDNEY DISEASE 1 GENE AND USES THEREOF

(57) Abstract

Autosomal dominant polycystic kidney disease (ADPKD) is a common genetic disorder which frequently results in renal failure, due to progressive cyst development. The major locus, PKD1, maps to 16p13.3. A chromosome translocation is identified associated with ADPKD which disrupts a gene (PBP), encoding a 14 kb transcript, in the PKD1 candidate region. Further mutations of the PBP gene were found in PKD1 patients confirming that PBP is the PKD1 gene. This gene is located adjacent to the tuberous sclerosis (2) locus in a genomic region that is reiterated more proximally on 16p. The duplicate area encodes three transcripts substantially homologous to the PKD1 transcript. Partial sequence analysis of the PKD1 transcript shows that it encodes a novel protein. Screening of actual or suspected ADPKD patients for normal or mutated PKD1 can be used for diagnostic purposes. PKD1-associated disorders such as ADPKD may be treated or prevented by PKD1 gene therapy and/or administration of functional PKD1 protein to affected cells.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

POLYCYSTIC KIDNEY DISEASE 1 GENE AND USES THEREOF

The present invention relates to the polycystic kidney disease 1 (PKD1) gene, mutations thereof in patients having PKD1-associated disorders, the protein encoded by the PKD1 gene, and their uses in diagnosis and therapy.

Background to the Invention

All references mentioned herebelow are listed in full at the end of the description which are herein incorporated by reference in their entirety. Except where the context clearly indicates otherwise, references to the PBP gene, transcript, sequence, protein or the like can be read as referring to the PKD1 gene, transcript, sequence, protein or the like, respectively.

A landmark study by Dalgaard, 1957 showed that autosomal dominant polycystic kidney disease (ADPKD) also termed adult polycystic kidney disease (APKD) is one of the commonest genetic diseases of man (approximately 1/1000 individuals affected). The major feature of this dominant disease is the development of cystic kidneys which commonly leads to renal failure in adult life. This simple description, however, belies the diverse systemic disorder, affecting many other organs (reviewed in Gabow, 1990) and one which occasionally presents in childhood (Fink, et al., 1993; Zerres, et al., 1993). Extrarenal manifestations include liver cysts (Milutinovic, et al., 1980), and more rarely cysts of the pancreas (Gabow, 1993) and other organs. Intracranial aneurysms occur in approximately 5% of patients and are a significant cause of morbidity and mortality due to subarachnoid haemorrhage (Chapman, et al., 1992). More recently, an increased prevalence of cardiac valve defects (Hossack, et al., 1988), herniae (Gabow, 1990) and colonic

diverticulae (Scheff, et al., 1980) has been reported.

The major cause of morbidity in ADPKD, however, is progressive renal disease characterised by the formation and enlargement of fluid filled cysts, resulting in grossly enlarged kidneys. Renal function deteriorates as normal tissue is compromised by cystic growth, resulting in end stage renal disease (ESRD) in more than 50% of patients by the age of 60 years (Gabow, et al., 1992): ADPKD accounts for 8-10% of all renal transplantation and dialysis patients in Europe and the USA (Gabow, 1993). Biochemical studies have suggested several potential causes of cyst formation and development, including: abnormal epithelial cell growth, alterations to the extracellular matrix and changes in cellular polarity and secretion (reviewed in Gabow, 1991; Wilson and Sherwood, 1991). The primary defect in ADPKD, however, remains unclear and considerable effort has therefore been applied to identifying the defective gene(s) in this disorder by genetic approaches.

The first step towards positional cloning of an ADPKD gene was the demonstration of linkage of one locus now designated the polycystic kidney disease 1 (PKD1) locus to the a globin cluster on the short arm of chromosome 16 (Reeders, et al., 1985). Subsequently, families with ADPKD unlinked to markers of 16p were described (Kimberling, et al., 1988; Romeo, et al., 1988) and a second ADPKD locus (PKD2) has recently been assigned to chromosome region 4q13-q23 (Kimberling, et al., 1993; Peters, et al., 1993). It is estimated that approximately 85% of ADPKD is due to PKD1 (Peters and Sandkuijl, 1992) with PKD2 accounting for most of the remainder. PKD2 appears to be a milder condition with a later age of onset and ESRD (Parfrey, et al., 1990; Gabow, et al., 1992; Ravine, et al., 1992).

The position of the PKD1 locus was refined to chromosome band 16p13.3 and many markers were isolated from that region (Breuning, et al., 1987; Reeders, et al., 1988; Breuning, et al., 1990; Germino, et al., 1990; Hyland, et al., 1990; Himmelbauer, et al., 1991). Their order, and the position of the PKD1 locus, has been determined by extensive linkage analysis in normal and PKD1 families and by the use of a panel of somatic cell hybrids (Reeders, et al., 1988; Breuning, et al., 1990; Germino, et al., 1990). An accurate long range restriction map (Harris, et al., 1990; Germino, et al., 1992) has located the PKD1 locus in an interval of approximately 600 kb between the markers GGG1 and SM7 (Harris, et al., 1991; Somlo, et al., 1992) (see Figure 1a). The density of CpG islands and identification of many mRNA transcripts indicated that this area is rich in gene sequences. Germino et al (1992) estimated that the candidate region contains approximately 20 genes.

Identification of the PKD1 gene from within this area has thus proved difficult and other means to pinpoint the disease gene were sought. Linkage disequilibrium has been demonstrated between PKD1 and the proximal marker VK5, in a Scottish population (Pound, et al., 1992) and between PKD1 and BLu24 (see Figure 1a), in a Spanish population (Peral, et al., 1994). Studies with additional markers have shown evidence of a common ancestor in a proportion of each population (Peral, et al., 1994; Snarey, et al., 1994), but the association has not precisely positioned the PKD1 locus.

Disease associated genomic rearrangements, detected by cytogenetics or pulsed field gel electrophoresis (PFGE) have been instrumental in the identification of various genes associated with various genetic disorders. Hitherto, no such abnormalities

related to PKD1 have been described. This situation contrasts with that for the tuberous sclerosis locus, which lies within 16p13.3 (TSC2). In that case, TSC associated deletions were detected by PFGE within the interval thought to contain the PKD1 gene and their characterisation was a significant step toward the rapid identification of the TSC2 gene (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). The TSC2 gene therefore maps within the candidate region for the hitherto unidentified PKD1 gene; as polycystic kidneys are a feature common to TSC and ADPKD1 (Bernstein and Robbins, 1991) the possibility of an aetiological link, as proposed by Kandt et al. (1992), was considered.

We have now identified a pedigree in which the two distinct phenotypes, typical ADPKD or TSC, are seen in different members. In this family, the two individuals with ADPKD are carriers of a balanced chromosome translocation with a breakpoint within 16p13.3. We have located the chromosome 16 translocation breakpoint and a gene disrupted by this rearrangement has been defined; the discovery of additional mutations of that gene in other PKD1 patients shows that we have identified the PKD1 gene.

Summary of the Invention

Accordingly, in one aspect, this invention provides an isolated, purified or recombinant nucleic acid sequence comprising:-

(a) a PKD1 gene or its complementary strand,

(b) a sequence substantially homologous to, or capable of hybridising to, a substantial portion of a molecule defined in (a) above,

(c) a fragment of a molecule defined in (a) or (b) above. In particular, there is provided a sequence wherein the PKD1 gene has the partial nucleic acid sequence according to Figure 7 and/or 10. The

invention therefore includes a DNA molecule selected from:

(a) a PKD1 gene or its complementary strand,

(b) a sequence substantially homologous to, or capable of hybridising to, a substantial portion of a molecule defined in (a) above,

(c) a molecule coding for a polypeptide having the partial sequence of Figure 7,

(d) genomic DNA corresponding to a molecule in (a) above; and

(e) a fragment of a molecule defined in any of (a), (b), (c) or (d) above.

The PKD1 gene described herein is a gene found on human chromosone 16, and the results of familial studies described herein form the basis for concluding that this PKD1 gene encodes a protein called PKD1 protein which has a role in the prevention or suppression of ADPKD. The PKD1 gene therefore includes the DNA sequences shown in Figures 7 and 10, and all functional equivalents. The gene furthermore includes regulatory regions which control the expression of the PKD1 coding sequence, including promotor, enhancer and terminator regions. Other DNA sequences such as introns spliced from the end-product PKD1 RNA transcript are also encompassed. Although work has been carried out in relation to the human gene, the corresponding genetic and functional sequences present in lower animals are also encompassed.

The present invention therefore further provides a PKD1 gene or its complementary strand having the partial sequence according to Figure 7. In particular, it provides a PKD1 gene or its complementary strand having the partial sequence of Figures 7 and/or 10 which gene or strand is mutated in some ADPKD patients (more specifically, PKD1 patients).

The invention further provides a nucleic acid sequence comprising a mutant PKD1 gene, especially one selected from a sequence comprising a partial sequence according to Figures 7 and/or 10 when:

(a) [OX114] base pairs 1746-2192 as defined in Figure 7 are deleted (446bp);

(b) [OX32] base pairs 3696-3831 as defined in Figure 7 are deleted by a splicing defect;

(c) [OX875] about 5.5kb flanked by the two Xba1 sites shown in Figure 3a are deleted and the EcoR1 site separating the CW10 (41kb) and JH1 (18kb) sites is thereby absent

(d) [WS53] about 100kb extending between the JH1 and CW21 and the SM6 and JH17 sites shown in Figure 6 and the PKD1 gene is thereby absent, the deletion lying proximally between SM6 and JH13;

(e) [461] 18bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(f) [OX1054] 20bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(g) [WS212] about 75kb are deleted between SM9-CW9 distally and the PKD1 3'UTR proximally as shown in Figure 12;

(h) [WS-215] about 160kb are deleted between CW20 and SM6-JH17 as shown in Figure 12;

(i) [WS-227] about 50kb are deleted between CW20 and JH11 as shown in Figure 12;

(j) [WS-219] about 27kb are deleted between JH1 and JH6 as shown in Figure 12;

(k) [WS-250] about 160kb are deleted between CW20 and BLu24 as shown in Figure 12;

(l) [WS-194] about 65kb is deleted between CW20 and CW10.

The invention therefore extends to RNA molecules comprising an RNA sequence corresponding to any of the DNA sequences set out above. The molecule is preferably the transcript reference PBP and

identifiable from the restriction map of Figure 3a and having a sequence of about 14 Kb.

In another aspect, the invention provides a nucleic acid probe having a sequence as set out above; in particular, this invention extends to a purified nucleic acid probe which hybridises to at least a portion of the DNA or RNA molecule of any of the preceding sequences. Preferably, the probe includes a label such as a radiolabel for example a ^{32}P label.

In another aspect, this invention provides a purified DNA or RNA coding for a protein comprising the amino acid sequence of Figure 7 and/or 10, or a protein polypeptide having homologous properties with said protein, or having at least one functional domain or active site in common with said protein.

The DNA molecule defined above may be incorporated in a recombinant cloning vector for expressing a protein having the amino acid sequence of Figure 7 and/or 10, or a protein or a polypeptide having at least one functional domain or active site in common with said protein.

In another aspect, the invention provides a polypeptide encoded by a sequence as set out above, or having the amino acid sequence according to the partial amino acid sequence of Figure 7 and/or 10, or a protein or polypeptide having homologous properties with said protein, or having at least one functional domain or active site in common with said protein. In particular, there is provided an isolated, purified or recombinant polypeptide comprising a PKD1 protein or a mutant or variant thereof or encoded by a sequence set out above or a variant thereof having substantially the same activity as the PKD1 protein.

This invention also provides an <u>in vitro</u> method of determining whether an individual is likely to be

affected with tuberous sclerosis, comprising the steps of:

assaying a sample from the individual to determine the presence and/or amount of PKD1 protein or polypeptide having the amino acid sequence of Figure 7 and/or 10.

Additionally or alternatively, a sample may be assayed to determine the presence and/or amount of mRNA coding for the protein or polypeptide having the amino acid sequence of Figure 7 and/or 10, or to determine the fragment lengths of fragments of nucleotide sequences coding for the protein or polypeptide of Figure 7 and/or 10, or to detect inactivating mutations in DNA coding for a protein having the amino acid sequence of Figure 7 and/or 10 or a protein having homologous properties. Said screening preferably includes applying a nucleic acid amplification process to said sample to amplify a fragment of the DNA sequence. Said nucleic acid amplification process advantagously utilizes at least one of the following sets of primers as identified herein:-

AH3 F9 : AH3 B7
3A3 C1 : 3A3 C2
AH4 F2 : JH14 B3

Alternatively, said screening method may comprise digesting said sample to provide EcoRI fragments and hybridising with a DNA probe which hybridises to the EcoRI fragment identified (A) in Figure 3(a), and said DNA probe may comprise the DNA probe CW10 identified herein.

Another screening method may comprise digesting said sample to provide BamHI fragments and hybridising with a DNA probe which hybridises to the BamHI fragment

identified (B) in Figure 3 (a), and said DNA probe may comprise the DNA probe 1A1H.6 identified herein.

A method according to the present invention may comprise detecting a PKD1-associated disorder in a patient suspected of having or having predisposition to, said disorder, the method comprising detecting the presence of and/or evaluating the characteristics of PKD1 DNA, PKD1 mRNA and/or PKD1 protein in a sample taken from the patient. Such method may comprise detecting and/or evaluating whether the PKD1 DNA is deleted, missing, mutated, aberrant or not expressing normal PKD1 protein. One way of carrying out such a method comprises:

A. taking a biological, tissue or biopsy sample from the patient;

B. detecting the presence of and/or evaluating the characteristics of PKD1 DNA, PKD1 mRNA and/or PKD1 protein in the sample to obtain a first set of results;

C. comparing the first set of results with a second set of results obtained using the same or similar methodology for an individual <u>not</u> suspected of having said disorders; and if the first and second sets of results differ in that the PKD1 DNA is deleted, missing, aberrant, mutated or not expressing PKD1 protein then that indicates the presence, predisposition or tendency of the patient to develop said disorders.

A specific method according to the invention comprises extracting a sample of PKD1 DNA or DNA from the PKD1 locus purporting to be PKD1 DNA from a patient, cultivating the sample <u>in vitro</u> and analysing the resulting protein, and comparing the resulting protein with normal PKD1 protein according to the well-established Protein Truncation Test.

Less sensitive tests include analysis of RNA using RT PCR (reverse transcriptase polymerase chain

reaction) and examination of genomic DNA.

On the other hand, if step C of the method is replaced by:

C. comparing the first set of results with a second set of results obtained using the same or similar methodology in an individual known to have the or at least one of said disorder(s); and if the first and second sets of results are substantially identical, this indicates that the PKD1 DNA in the patient is deleted, mutated or not expressing normal PKD1 protein.

The invention further provides a method of characterising a mutation in a subject suspected of having a mutation in the PKD1 gene, which method comprises:

A. amplifying each of the exons in the PKD1 gene of the subject;

B. denaturing the complementary strands of the amplified exons;

C. diluting the denatured separate, complementary strands to allow each single-stranded DNA molecule to assume a secondary structural conformation;

D. subjecting the DNA molecule to electrophoresis under non-denaturing conditions;

E. comparing the electrophoresis pattern of the single-stranded molecule with the electrophoresis pattern of a single-stranded molecule containing the same amplified exon from a control individual which has either a normal or PKD1 heterozygous genotype; and

F. sequencing any amplification product which has an electrophoretic pattern different from the pattern obtained from the DNA of the control individual.

The invention also extends to a diagnostic kit for carrying out a method as set out above, comprising nucleic acid primers for amplifying a fragment of the DNA or RNA sequences defined above. The nucleic acid

primers may comprise at least one of the following sets:

AH3 F9 : AH3 B7
3A3 C1 : 3A3 C2
AH4 F2 : JH14 B3

Another embodiment of kit may combine one or more substances for digesting a sample to provide EcoRI fragments and a DNA probe as previously defined.

A further embodiment of kit may comprise one or more substances for digesting a sample to provide BamHI fragments and a DNA probe as previously defined.

Still further, a kit may include a nucleic acid probe capable of hybridising to the DNA or RNA molecule previously defined.

A vector (such as Bluscript (available from Stratagene)) comprising a nucleic acid sequence set out above; and a host cell (such as E. coli strain SL-1 Blue (available from Stratagene)) transfected or transformed with the vector are also provided, together with the use of such a vector or a nucleic acid sequence set out above in gene therapy and/or in the preparation of an agent for treating or preventing a PKD1-associated disorder. Therefore there is further provided a method of treating or preventing a PKD1-associated disorder which method comprises administering to a patient in need thereof a functional PKD1 gene to affected cells in a manner that permits expression of PKD1 protein therein and/or a transcript produced from a mutated chromosome (such as the deleted WS-212 chromosome) which is capable of expressing functional PKD1 protein therein.

The invention also extends to any inventive combination of features set out above or in the following description.

Brief Description Of The Drawings

Figure 1a (top): A long range map of the terminal region of the short arm of chromosome 16 showing the PKD1 candidate region defined by genetic linkage analysis. The positions of selected DNA probes and microsatellites used for haplotype, lindage or heterozygosity analyses are indicated. Markers previously described in linkage disequilibrium studies are shown in bold (from: Harris, et al., 1990; Harris, et al., 1991; Germino, et al., 1992; Somlo, et al., 1992; Peral, et al., 1994; Snarey, et al., 1994).

(bottom): A detailed map of the distal part of the PKD1 candidate region showing: the area of 16p13.3 duplicated in 16p13.1 (hatched); C, Cla I restriction sites; the breakpoints in the somatic cell hybrids, N-OH1 and P-MWH2A; DNA probes and the TSC2 gene. The limits of the position of the translocation breakpoint found in family 77 (see b), determined by evidence of heterozygosity (in 77-4) and PFGE (see c and text) is also indicated. The contig covering the 77 breakpoint region consists of the cosmids: 1, CW9D; 2, ZDS5; 3, JH2A; 4, REP59; 5, JC10.2B; 6, CW10III; 7, SM25A; 8, SMII; 9, NM17.

Figure 1b: Pedigree of family 77 which segregates a 16;22 translocation; showing the chromosomal composition of each subject. Individuals 77-2 and 77-3 have the balanced products of the exchange - and have PKD1; 77-4 is monosomic for 16p13.3-->16pter and 22q11.21-->22pter - and has TSC.

Figure 1c: PFGE of DNA from members of the 77 family: 77-1 (1); 77-2 (2); 77-3 (3); 77-4 (4); digested with Cla I and hybridised with SM6. In addition to the normal fragments of 340 and partially digested fragment of 480 kb a proximal breakpoint fragment of approximately 100 kb (arrowed) is seen in individuals, 77-2, 77-3 and 77-4; concordant with

segregation of the der(16) chromosome.

Figure 2: FISH of the cosmid CW10III (cosmid 6; Figure 1a) to a normal male metaphase. Duplication of this locus is illustrated with two sites of hybridisation on 16p; the distal site (the PKD1 region) is arrowed. The signal from the proximal site (16p13.1) is stronger than that from the distal, indicating that sequences homologous to CW10III are re-iterated in 16p13.1.

Figure 3a: A detailed map of the 77 translocation region showing the precise localisation of the 77 breakpoint and the region that is duplicated in 16p13.1 (hatched). DNA probes (open boxes); the transcripts, PKD1 and TSC2 (filled boxes; with direction of transcription indicated by an arrow) and cDNAs (grey boxes) are shown below the genomic map. The known genomic extent of each gene is indicated at the bottom of the diagram and the approximate genomic locations of each cDNA is indicated under the genomic map. The positions of genomic deletions found in PKD1 patients, OX875 and OX114, are also indicated. Restriction sites for EcoR I (E) and incomplete maps for BamH I (B); Sac I (S) and Xba I (X) are shown. SM3 is a 2kb BamH1 fragment shown at the 5' end of the gene.

Figure 3b: Southern blots of BamH I digested DNA from individuals: 77-1 (1); 77-2 (2); and 77-4 (4) hybridised with: left panel, 8S3 and right panel, 8S1 (see a). 8S3 detects a novel fragment on the telomeric side of the breakpoint (12 kb: arrowed) associated with the der(22) chromosome in 77-2, but not 77-4; 8S1 identifies a novel fragment on the centromeric side of the breakpoint (9 kb: arrowed) - associated with the der(16) chromosome - in 77-2 and 77-4. The telomeric breakpoint fragment is also seen weakly with 8S1 (arrowed) indicating that the breakpoint lies in the distal part of 8S1. The 8S3 and 8S1 loci are both

duplicated; the normal BamH I fragment detected at the 16p13.3 site by these probes is 11 kb (see a), but a similar sized fragment is also detected at the 16p13.1 site. Consequently, the breakpoint fragments are much fainter than the normal (16p13.1 plus 16p13.3) band.

Figure 4a: PBP cDNA, 3A3, hybridised to a Northern blot containing ~1 mg polyA selected mRNA per lane of the tissue specific cell lines: lane 1, MJ, EBV-transformed lymphocytes; lane 2, K562, erythroleukaemia; lane 3, FS1, normal fibroblasts; lane 4, HeLa, cervical carcinoma; lane 5, G401, renal Wilm's tumour; lane 6, Hep3B, hepatoma; lane 7, HT29, colonic adenocarcinoma; lane 8, SW13, adrenal carcinoma; lane 9, G-CCM, astrocytoma. A single transcript of approximately 14 kb is seen; the highest level of expression is in fibroblasts and in the astrocytoma cell line, G-CCM. Although in this comparative experiment little expression is seen in lanes 1, 4 and 7, we have demonstrated at least a low level of expression in these cell lines on other Northern blots and by RT-PCR (see later).

Figure 4b: A Northern blot containing ~ 20 mg of total RNA from the cell line G-CCM hybridised with cDNAs or a genomic probe which identify various parts of the PBP gene. Left panel, a single ~14 kb transcript is seen with a cDNA from the single copy area, 3A3. Right panel, a cDNA, 21P.9, that is homologous to parts of the region that is duplicated (JH12, JH8 and JH10; see Figure 3a) hybridises to the PBP transcript and three novel transcripts; HG-A (~ 21 kb), HG-B (~ 17 kb) and HG-C (8.5 kb). A similar pattern of transcripts is seen with cDNAs and genomic fragments that hybridise to the area between JH5 and JH13, with the exception of the JH8 area. Middle panel, JH8 hybridises to the transcripts PBP, HG-A and HG-B but not to HG-C.

Figure 4c: A Northern blot of 20mg total fibroblast RNA from: normal control (N); 77-2 (2); 77-4 (4) hybridised with 8S1, which contains the 16;22 translocation breakpoint (see Figure 3). A transcript of ~ 9 kb (PBP-77) is identified in the two patients with this translocation but not in the normal control. PBP-77 is a chimeric PBP transcript formed due to the translocation and is not seen in 77-2 or 77-4 RNA with probes which map distal to the breakpoint.

Figure 5a: FIGE of DNA from: normal (N) and ADPKD patient OX875 (875), digested with EcoR I and hybridised with, left panel, CW10; middle panel, JH1. Normal fragments of 41 kb (plus a 31 kb fragment from the 16p13.1 site), CW10, and 18 kb, JHI, are identified with these probes; OX875 has an additional 53 kb band (arrowed). The EcoR I site separating these two fragments is removed by the deletion (see Figure 3a). The right panel shows a Southern blot of BamH I digested DNA (as above) hybridised with 1A1H.6. A novel fragment of 9.5 kb is seen in OX875 DNA, as well as the normal 15 kb fragment. These results indicate that OX875 has a 5.5 kb deletion; its position was determined more precisely by mapping relative to two Xba I sites which flank the deletion (see figure 3a).

Figure 5b: Northern blot of total fibroblast RNA, as (a), hybridised with the cDNAs, AH4, 3A3 and AH3. A novel transcript (PBP-875) of ~ 11 kb is seen with AH4 (the band is reduced in intensity because the probe is partly deleted) and AH3 (arrowed), which flank the deletion, but not 3A3 which is entirely deleted (see figure 3a). The transcripts HG-A, HG-B and HG-C, from the duplicated area, are seen with AH3 (see figure 4b).

Figure 5c: Left panel; FIGE of DNA from: normal (N) and ADPKD patient OX114 (114), digested with EcoR I and hybridised with CW10; a novel fragment of 39 kb (arrowed) is seen in OX114. Middle panel; DNA, as

above, plus the normal mother (M) and brother (B) of OX114 digested with BamH I and hybridised with CW21. A larger than normal fragment of 19 kb (arrowed) was detected in OX114 but not other family members due to deletion of a BamH I site; together these results are consistent with a 2 kb deletion (see Figure 3a). Right panel; RT-PCR of RNA, as above, with primers flanking the OX114 deletion (see Experimental Procedures). A novel fragment of 810 bp (arrowed) is seen in OX114, indicating a deletion of 446 bp in the PBP transcript.

Figure 5d: RT-PCR of RNA from: ADPKD patient OX32 (32) plus the probands, normal mother (M) and affected father (F) and sibs (1) and (2) using the C primer pair from 3A3 (see Experimental Procedures). A novel fragment of 125 bp is detected in each of the affected individuals.

Figure 6: Map of the region containing the TSC2 and PBP genes showing the area deleted in patient WS-53 and the position of the 77 translocation breakpoint. Localisation of the distal end of the WS-53 deletion was previously described (European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and we have now localised the proximal end between SM6 and JHI7. The size of the aberrant Mlu I fragment in WS-53, detected by JH1 and JH17, is 90kb and these probes lie on adjacent Mlu I fragments of 120kb and 70kb, respectively. Therefore the WS-53 deletion is ~ 100kb. Restriction sites for: Mlu I (M); Nru I (R); Not I (N); and partial maps for Sac II (S) and BssH II (H) are shown. DNA probes (open boxes) and the TSC2 and PBP transcripts (filled boxes) are indicated below the line with their known genomic extents (brackets). The locations of the microsatellites KG8 and SM6 are also indicated.

Figure 7: The partial nucleotide sequence (cDNA) of the PKD1 transcript extending 5631bp to the 3' end

of the gene. The corresponding predicted protein (also shown in SEQ ID NO: 4:) is shown below the sequence and extends from the start of the nucleotide sequence. The GT-repeat, KG8, is in the 3' untranslated region between 5430-5448 bp. This sequence corresponds to GenBank Accession No. L33243 and is shown in SEQ ID NO: 3:.

Figure 8: The sequence of the probe 1A1H0.6 (also shown in SEQ ID NO: 5:).

Figure 9: The sequence (SEQ ID NO: 6:) of the probe CW10 which is about 0.5kb.

Figure 10: The larger partial nucleotide sequence (SEQ ID NO: 1:) of the PKD1 transcript (cDNA) extending from bp 2 to 13807bp to the 3' end of the gene together with the corresponding predicted protein (also shown in SEQ ID NO: 2:). This larger partial sequence encompasses the (smaller) partial sequence of Figure 7 from amino acid no. 2726 in SEQ ID NO: 3: and relates to the entire PKD1 gene sequence apart from its extreme 5' end.

Figure 11: A map of the 75bp intron amplified by the primer set 3A3C insert at position 3696 of the 3' sequence showing the positions of genomic deletions found in PKD1 patients 461 and OX1054.

Figure 12: A map of the region of chromosome 16 containing the TSC2 and PKD1 genes showing the areas affected in patients WS-215, WS-250, WS-212, WS-194, WS-227 and WS-219; also WS-53 (but cf. Figure 6). Genomic sites for the enzymes MluI (M), ClaI (C), PvuI (P) and NruI (R) are shown. Positions of single copy probes and cosmids used to screen for deletions are shown below the line which represents ~400kb of genomic DNA. The genomic distribution of the approximately 45kb TSC2 gene and known extent of the PKD1 gene are indicated above. The hatched area respresents an ~50kb

region which is duplicated more proximally on chromosome 16p.

Detailed Description of the Drawings

A translocation associated with ADPKD

A major pointer to the identity of the PKD1 gene was provided by a Portuguese pedigree (family 77) with both ADPKD and TSC (Figure 1b). Cytogenetic analysis showed that the mother, 77-2, has a balanced translocation, 46XX t(16;22)(p13.3;q11.21) which was inherited by her daughter, 77-3. The son, 77-4, has the unbalanced karyotype, 45XY-16-22+der(16)(16qter-->16p13.3: :22q11.21-->2qter) and consequently is monosomic for 16p13.3-->16pter as well as for 22q11.21-->22pter. This individual has the clinical phenotype of TSC (see Experimental Procedures); the most likely explanation is that the TSC2 locus located within 16p13.3 is deleted in the unbalanced karyotype.

Further analysis revealed that the mother (77-2), and the daughter (77-3) with the balanced translocation, have the clinical features of ADPKD (see Experimental Procedures), while the parents of 77-2 were cytogenetically normal, with no clinical features of TSC and no renal cysts on ultrasound examination (aged 67 and 82 years). Although kidney cysts can be a feature of TSC, no other clinical signs of TSC were identified in 77-2 or 77-3, making it unlikely that the polycystic kidneys were due to TSC. We therefore investigated the possibility that the translocation disrupted the PKD1 locus in 16p13.3 and proceeded to identify and clone the region containing the breakpoint.

The 77 family was analysed with polymorphic markers from 16p13.3. Individual 77-4 was hemizygous for MS205.2 and GGG1, but heterozygous for SM6 and more proximal markers, locating the translocation breakpoint

between GGG1 and SM6 (see Figure 1a). Fluorescence in situ hybridisation (FISH) of a cosmid from the TSC2 region, CW9D (cosmid 1 in Figure 1a), to metaphase spreads showed that it hybridised to the der(22) chromosome of 77-2; placing the breakpoint proximal to CW9D and indicating that 77-4 was hemizygous for this region consistent with his TSC phenotype. DNA from members of the 77 family was digested with Cla I, separated by PFGE and hybridised with SM6; revealing a breakpoint fragment of ~ 100 kb in individuals with the der(16) chromosome (Figure 1c). The small size of this novel fragment enabled the breakpoint to be localised distal to SM6 in a region of just 60 kb (Figure 1a). A cosmid contig covering this region was therefore constructed (see Experimental Procedures for details).

The translocation breakpoint lies within a region duplicated elsewhere on chromosome 16p (16p13.1)

It was previously noted that the region between CW21 and N54 (Figure 1a) was duplicated at a more proximal site on the short arm of chromosome 16 (Germino, et al., 1992; European Chromosome 16 Tuberous Sclerosis Consortium, 1993). Figure 2 shows that a cosmid, CW10III, from the duplicated region hybridises to two points on 16p; the distal, PKD1 region and a proximal site positioned in 16p13.1. The structure of the duplicated area is complex with each fragment present once in 16p13.3 re-iterated two-four times in 16p13.1 (see Figure 2). Cosmids spanning the duplicated area in 16p13.3 were subcloned (see Figure 3a and Experimental Procedures for details) and a restriction map was generated. A genomic map of the PKD1 region was constructed using a radiation hybrid, Hy145.19 which contains the distal portion of 16p but not the duplicate site in 16p13.1.

To localise the 77 translocation breakpoint, subclones from the target region were hybridised to 77-

2 DNA, digested with Cla I and separated by PFGE. Once probes mapping across the breakpoint were identified they were hybridised to conventional Southern blots of 77 family DNA. Figure 3b shows that novel BamH I fragments were detected from the centromeric and telomeric side of the breakpoint, which was localised to the distal part of the probe 8S1 (Figure 3a). Hence, the balanced translocation was not associated with a substantial deletion, and the breakpoint was located more than 20 kb proximal to the TSC2 locus (Figure 3a). These results supported the hypothesis that polycystic kidney disease in individuals with the balanced translocation (77-2 and 77-3) was not due to disruption of the TSC2 gene, but indicated that a separate gene mapping just proximal to TSC2, was likely to be the PKD1 gene.

The polycystic breakpoint (PBP) gene is disrupted by the translocation

Localisation of the 77 breakpoint identified a precise region in which to look for a candidate for the PKD1 gene. During the search for the TSC2 gene we identified other transcripts not associated with TSC including a large transcript (~ 14 kb) partially represented in the cDNAs 3A3 and AH4 which mapped to the genomic fragments CW23 and CW21 (Figure 3a). The orientation of the gene encoding this transcript had been determined by the identification of a polyA tract in the cDNA, AH4: the 3' end of this gene lies very close to the TSC gene, in a tail to tail orientation (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). To determine whether this gene crossed the translocation breakpoint genomic probes from within the duplicated area and flanking the breakpoint were hybridised to Northern blots. Probes from both sides of the breakpoint, between JH5 and JH13 identified the 14 kb transcript (Figure 3a and see below for details).

Therefore, this gene previously called 3A3, but now designated the PBP gene extended over the 77 breakpoint and consequently was a candidate for the PKD1 gene. A walk was initiated to increase the extent of the PBP cDNA contig and several new cDNAs were identified using probes from the single copy (non-duplicated) region (see Experimental Procedures for details). A cDNA contig was constructed which extended ~5.7 kb, including ~2 kb into the area that is duplicated (Figure 3a).

Expression of the PBP gene

Initial studies of the expression pattern of the PBP gene were undertaken with cDNAs that map entirely within the single copy region (e.g. AH4 and 3A3). Figure 4a shows that the ~ 14 kb transcript was identified by 3A3 in various tissue-specific cell lines. From this and other Northern blots we concluded that the PBP gene was expressed in all of the cell lines tested, although often at a low level. The two cell lines which showed the highest level of expression were fibroblasts and a cell line derived from an astrocytoma, G-CCM. Significant levels of expression were also obtained in cell lines derived from kidney (G401) and liver (Hep3B). Measuring the expression of the PBP gene in tissue samples by Northern blotting proved difficult because such a large transcript is susceptable to minor RNA degradation. However, initial results with an RNAse protection assay, using a region of the gene located in the single copy area (see Experimental Procedures), showed a moderate level of expression of the PBP gene in tissue obtained from normal and polycystic kidney (data not shown). The widespread expression of the PBP gene is consistent with the systemic nature of ADPKD.

Identification of transcripts that are partially homologous to the PBP transcript

New cDNAs were identified with the genomic fragments, JH4 and JH8, that map to the duplicated region (Figure 3a and see Experimental Procedures). However, when these cDNAs were hybridised to Northern blots a more complex pattern than that seen with 3A3 was observed. As well as the ~14 kb PBP transcript, three other, partially homologous transcripts were identified designated homologous gene-A (HG-A; ~ 21 kb), HG-B (~ 17 kb) and HG-C (8.5 kb) (Figure 4b). There were two possible explanations for these results, either the HG transcripts were alternatively spliced forms of the PBP gene, or the HG transcripts were encoded by genes located in 16p13.1. To determine the genomic location of the HG loci a fragment from the 3' end of one HG cDNA (HG-4/1.1) was isolated. HG-4/1.1 hybridised to all three HG transcripts, but not to the PBP transcript and on a hybrid panel it mapped to 16p13.1 (not the PKD1 area). These results show that all the HG transcripts are related to each other outside the region of homology with the PBP transcript and that the HG loci map to the proximal site (16p13.1).

An abnormal transcript associated with the 77 translocation

As the PBP gene was transcribed across the region disrupted by the 77 translocation breakpoint, in a proximal to distal direction on the chromosome (see Figure 3a) it was possible that a novel transcript originating from the PBP promotor would be found in this family. Figure 4c shows that using a probe to the PBP transcript that mapped mainly proximal to the breakpoint, a novel transcript of approximately 9 kb (PBP-77) derived from the der(16) product of the translocation was detected. Interestingly, the PBP-77

transcript appears to be expressed at a higher level than the normal PBP product. These results confirmed that the 77 translocation disrupts the PBP gene and supports the hypothesis that this is the PKD1 gene.

Mutations of the PBP gene in other ADPKD patients

To prove that the PBP gene is the defective gene at the PKD1 locus, we analysed this region for mutations in patients with typical ADPKD. The 3' end of the PBP gene was most accessible to study as it maps outside the duplicated area. To screen this region BamH I digests of DNA from 282 apparently unrelated ADPKD patients were hybridised with the probe 1A1H.6, (see Figure 3a). In addition, a large EcoR I fragment (41 kb) which contains a significant proportion of the PBP gene was assayed by field inversion gel electrophoresis (FIGE) in 167 ADPKD patients, using the probe CW10. Two genomic rearrangements were identified in ADPKD patients by these procedures; each identified by both methods.

The first rearrangement was identified in patient OX875 (see Experimental Procedures for clinical details) who was shown to have a 5.5 kb genomic deletion within the 3' end of the PBP gene, producing a smaller transcript (PBP-875) (see Figures 5a, b and 3a for details). This genomic deletion results in a ~3 kb internal deletion of the transcript with the ~500 bp adjacent to the polyA tail intact. In this family linkage of ADPKD to chromosome 16 could not be proven because although OX875 has a positive family history of ADPKD there were no living, affected relatives. However, paraffin-embedded tissue from her affected father (now deceased) was available. We demonstrated that this individual had the same rearrangement as OX875 by PCR amplification of a 220bp fragment spanning the deletion (data not shown). This result and analysis of two unaffected sibs of OX875, that did not

have the deletion, showed that this mutation was transmitted with ADPKD.

The second rearrangement detected by hybridisation was a 2 kb genomic deletion within the PBP gene, in ADPKD patient OX114 (see Experimental Procedures for clinical details and Figures 5c and 3a). No abnormal PBP transcript was identified by Northern blot analysis, but using primers flanking the deletion (see Experimental Procedures) a shortened product was detected by RT-PCR (Figure 5c). This was cloned and sequenced and shown to have a frame-shift deletion of 446 bp (between base pair 1746 and 2192 of the sequence shown in Figure 7). OX114 is the only member of the family with ADPKD (she has no children) and ultrasound analysis of her parents at age 78 (father) and 73 years old (mother) showed no evidence of renal cysts. Somatic cell hybrids were produced from OX114 and the deleted chromosome was found to be of paternal origin by haplotype analysis. The father of OX114 is now deceased but analysis of DNA from the brother of OX114 (OX984) with seven microsatellite markers from the PKD1 region (see Experimental Procedures) showed that he shares the same paternal chromosome, in the PKD1 region, as OX114. Renal ultrasound revealed no cysts in OX984 at age 53 and no deletion was detected by DNA analysis (Figure 5c). Hence, the deletion in OX114 is a de novo event associated with the development of ADPKD. Although it is not possible to show that the ADPKD is chromosome 16-linked, the location of the PBP gene indicates that this is a de novo PKD1 mutation.

To identify more PKD1 associated mutations, single copy regions of the PBP gene were analysed by RT-PCR using RNA isolated from lymphoblastoid cell lines established from ADPKD patients. cDNA from 48 unrelated patients was amplified with the primer pair 3A3 C (see Experimental Procedures) and the product of 260 bp was

analysed on an agarose gel. In one patient, OX32, an additional smaller product (125 bp) was identified, consistent with a deletion or splicing mutation. OX32 comes from a large family in which the disease can be traced through three generations. Analysis of RNA from two affected sibs of OX32 and his parents showed that the abnormal transcript segregates with PKD1 (Figure 5d).

Amplification of normal genomic DNA with the 3A3 C primers generates a product of 418 bp; sequencing showed that this region contains two small introns (5', 75 bp and 3', 83 bp) flanking a 135 bp exon. The product amplified from OX32 genomic DNA was normal in size, excluding a genomic deletion. However, heteroduplex analysis of that DNA revealed larger heteroduplex bands, consistent with a mutation within that genomic interval. The abnormal OX32, RT-PCR product was cloned and sequenced: this demonstrated that, although present in genomic DNA, the 135 bp exon was missing from the abnormal transcript. Sequencing of OX32 genomic DNA demonstrated a G-->C transition at +1 of the splice donor site following the 135 bp exon. This mutation was confirmed in all available affected family members by digesting amplified genomic DNA with the enzyme Bst NI: a site is destroyed by the base substitution. The splicing defect results in an in-frame deletion of 135 bp from the PBP transcript (3696 bp to 3831 bp of the sequence shown in Figure 7). Together, the three intragenic mutations confirm that the PBP gene is the defective gene at the PKD1 locus.

Deletions that disrupt the TSC2 and the PKD1 gene

We previously identified a deletion (WS-53) which disrupts the TSC2 gene and the PKD1 gene (European Chromosome 16 Tuberous Sclerosis Consortium, 1993), although its full proximal extent was not determined. Further study has shown that the deletion extends ~ 100

kb (see Figure 6 for details) and deletes most if not all of the PKD1 gene. This patient has TSC but also has unusually severe polycystic disease of the kidneys. Other patients with a similar phenotype have also been under investigation. Deletions involving both TSC2 and PKD1 were identified and characterised in six patients in whom TSC was associated with infantile polycystic kidney disease. As well as the deletion in WS-53, those in WS-215 and "S-250 also extended proximally well beyond the known distribution of PKD1 and probably delete the entire gene. The deletion in WS-194 extended over the known extend of PKD1, but not much further proximally, while the proximal breakpoints in WS-219 and WS-227 lay within PKD1 itself. Northern analysis of case WS-219 with probe JH8, which lies outside the deletion, showed a reduced level of the PKD1 transcript but no evidence of an abnormally sized transcript (data now shown). Analysis of samples from the clinically unaffected parents of patients WS-53, WS-215, WS-219, WS-227 and WS-250 showed the deletions in these patients to be de novo. The father of WS-194 was unavailable for study.

In a further case (WS-212), renal ultrasound showed no cysts at four years of age but a deletion was identified which removed the entire TSC2 gene and deleted an XbaI site which is located 42bp 5' to the polyadenylation signal of PKD1. To determine the precise position of the proximal breakpoint in PKD1, a 587bp probe from the 3' untranslated region (3'UTRP) was hybridised to XbaI digested DNA. A 15kb XbaL breakpoint fragment was detected with an approximately equal intensity to the normal fragment of 6kb, indicating that most of the PKD13'UTR was preserved on the mutant chromosome. Evidence that a PKD1 transcript is produced from the deleted chromosome in WS-212 was obtained by 3' rapid identification of cDNA ends (RACE)

with a novel, smaller product generated from WS-212 cNDA. Characterisation of this product showed that polyadenylation occurs 546bp 5' to the normal position, within the 3'UTR of PKD1 (231bp 3' to the stop codon at 5073bp of the described PKD1 sequence [14]). A transcript with an intact open reading frame is thus produced from the deleted WS-212 chromosome. It is likely that a functional PKD1 protein in produced from this transcript, explaining the lack of cystic disease in this patient. The sequence preceeding the novel site of polyA addition is: AGTCAGTAATTTATATGGTGTTAAAATGTG(A)n. Although not conforming precisely to the concensus of AATAAA, it is likely that part of this AT rich region acts as an alternative polyadenylation signal if, as in this case, the normal signal is deleted (a possible sequence is underlined).

The WS-212 deletion if 75kb between SM9-CW9 distally and the PKD1 3'UTR proximally. The WS-215 deletion is 160kb between CW15 and SM6-JH17. WS-194 has 65kb deleted between CW20 and CW10-CW36. WS-227 has a 50kb deletion between CW20 and JH11 and WS-219 has a 27kb deletion between JH1 and JH6. The distal end of the WS-250 deletion is in CW20 but the precise location of the proximal end is not known. However, the same breakpoint fragment of 320kb is seen with Pvul-digested DNA using probes on adjacent Pvul fragments, CE18 (which normally detects a 245kb fragment) and BLu24 (235kb). Hence this deletion can be estimated ~160kb. b. PFGE analysis of the deletion in WS-219. Mlul digested DNA from a normal control (N) and WS-219 probed with the clones H2, JH1, CW21 and CW10 which detect an ~130kb fragment in normal individuals. CW10 also detects a much smaller fragment from the duplicated region situated more proximally on 16p. A novel fragment of ~100kb is seen in WS-219 with

probes H2 and CW10 which flank the deletion in this patient. JH1 is partially deleted but detects the novel band weakly. The aberrant fragment is not detected by CW-21, which is deleted on the mutant chromosome. BamH1 digested DNA of normal control (N) and WS-219 separated by conventional gel electrophoresis and hybridised to probes JH1 and JH6 which flank the deletion. The same breakpoint fragment of ~3kb is seen with both probes, consistent with a deletion of ~27kb ending within the BamH1 fragments seen by these probes.

Two further deletions

In addition we have characterised two further mutations of this gene which were identified in typical PKD1 families. In both cases the mutation is a deletion in the 75bp intron amplified by the primer pair 3A3C (European Polycystic Kidney Disease Consortium, 1994). The deletions are of 18bp and 20bp, respectively, in the patients 461 and OX1054. Although these deletions do not disrupt the highly conserved sequences flanking the exon/intron boundaries, they do result in aberrant splicing of the transcript. In both cases, two abnormal mRNAs are produced, one larger and one smaller than normal. Sequencing of these cDNAs showed that the larger transcript includes the deleted intron, and so has an in-frame insertion of 57bp in 461, while OX1054 has a frameshift insertion of 55bp. The smaller transcript is due to activation of a cryptic splice site in the exon preceding the deleted intron and results in an in-frame deletion of 66bp in both patients. The demonstration of two additional mutations of this gene in PKD1 patients further confirms that this is the PKD1 gene.

Characterisation of the PKD1 gene

To characterise the PKD1 gene further, evolutionary conservation was analysed by zoo

blotting'. Using probes from the single copy, 3' region (3A3) and from the duplicated area (JH4, JH8) the PKD1 gene was conserved in other mammalian species, including horse, dog, pig and rodents (data not shown). No evidence of related sequences were seen in chicken, frog or drosophila by hybridisation at normal stringency. The degree of conservation was similar when probes from the single copy or the duplicated region were employed.

The full genomic extent of the PKD1 gene is not yet known, although results obtained by hybridisation to Northern blots show that it extends from at least as far as JH13. Several CpG islands have been localised 5' of the known extent of the PKD1 gene (Figure 6), although there is no direct evidence that any of these are associated with this gene.

The cDNA contig extending 5631 bp to the 3' end of the PKD1 transcript was sequenced; where possible more than one cDNA was analysed and in all regions both strands were sequenced (Figure 7). We estimated that this accounts for ~40% of the PKD1 transcript. An open reading frame was detected which runs from the 5' end of the region sequenced and spans 4842 bp, leaving a 3' untranslated region of 789 bp which contains the previously described microsatellite, KG8 (Peral, et al., 1994; Snarey, et al., 1994). A polyadenylation signal is present at nucleotides 5598-5603 and a polyA tail was detected in two independent cDNAs (AH4 and AH6) at position, 5620. Comparison with the cDNAs HG-4 and 11BHS21, which are encoded by genes in the duplicate,16p13.1 region, show that 1866 bp at the 5' end of the partial PKD1 sequence shown in Figure 7 lies within the duplicated area. The predicted amino acid sequence from the available open reading frame extends 1614 residues, and is shown in Figure 7. A search of the swissprot and NBRF data bases with the available

protein sequence, using the Blast programme (Altschul, et al., 1990) identified only short regions of similarity (notably, between amino-acids 690-770 and 1390-1530) to a diverse group of proteins; no highly significant areas of homology were recognised. The importance of the short regions of similarity is unclear as the search for protein motifs with the ProSite Programme did not identify any recognised functional protein domains within the PKD1 gene.

The task of identifying and characterising the PKD1 gene has been more difficult than for other disorders because more than three quarters of the gene is embedded in a region of DNA that is duplicated elsewhere on chromosome 16. This segment of 40-50 kb of DNA, present as a single copy in the PKD1 area (16p13.3), is re-iterated as several divergent copies in the more proximal region, 16p13.1. This proximal site contains three gene loci (HG-A, -B and -C) that each produce polyadenylated mRNAs and share substantial homology to the PKD1 gene; it is not known whether these partially homologous transcripts are translated into functional proteins.

Although gene amplification is known as a major mechanism for creating protein diversity during evolution, the discovery of a human disease locus embedded within an area duplicated relatively recently is a new observation. In this case because of the recent nature of the reiteration the whole duplicated genomic region retains a high level of homology, not just the exons. The sequence of events leading to the duplication and which sequence represents the original gene locus are not yet clear. However, early evidence of homology of the 3' ends of the three HG transcripts which are different from the 3' end of the PKD1 gene indicated that the loci in 16p13.1 have probably arisen

by further reiteration of sequences at this site, after it separated from the distal locus.

To try to overcome the duplication problem we have employed an exon linking approach using RNA isolated from a radiation hybrid, Hy145.19, that contains just the PKD1 part of chromosome 16, and not the duplicate site in 16p13.1. Hence, this hybrid produces transcripts from the PKD1 gene but not from the homologous genes (HG-A, HG-B and HG-C). We have also sequenced much of the genomic region containing the PKD1 gene, from the cosmid JH2A, and have sequenced a number of cDNAs from the HG locus. To determine the likely position of PKD1 exons in the genomic DNA we compared HG cDNAs, (HG-4 and HG-7) to the genomic sequence. We then designed primers with sequences corresponding to the genomic DNA, to regions identified by the HG exons and employing cDNA generated from the hybrid Hy145.19, we amplified sections of the PKD1 transcript. The polymerase Pfu was used to minimise incorporation errors. These amplified fragments were then cloned and sequenced. The PDK1 cDNA contig whose sequence is shown in Figure 10 is made up of (3'-5') the original 5.7 kb of sequence shown in Figure 7, and the cDNAs: gap α 22 (890 bp), gap gamma (872 bp), a section of genomic DNA from the clone JH8 (2,724 bp) which corresponds to a large exon, S1-S3 (733 bp), S3-S4 (1,589 bp) and S4-S13 (1,372 bp). Together these make a cDNA of 13,807 bp with the extreme 5' end of the transcript still uncharacterised. When these cDNAs from the PKD1 contig were sequenced an open reading frame was found to run from the start of the contig to the previously-identified stop codon, a region of 13,018 bp. The predicted protein encoded by the PKD1 transcript is also shown in Figure 10 and has 4,339 amino acid residues.

We have therefore compelling evidence that mutations of the PKD1 gene give rise to the typical phenotype of ADPKD. The location of this gene within the PKD1 candidate region and the available genetic evidence from the families with mutations show that this is the PKD1 gene. The present invention therefore includes the PKD1 gene itself and the six PKD1-associated mutations which have been described: a de novo translocation, which was subsequently transmitted with the phenotype; two intragenic deletions (one a de novo event); two further deletions; and a splicing defect.

It has previously been argued that PKD1 could be recessive at the cellular level, with a second somatic mutation required to give rise to cystic epithelium (Reeders, 1992). This "two hit" process is thought to be the mutational mechanism giving rise to several dominant diseases, such as neurofibromatosis (Legius, et al., 1993) and tuberous sclerosis (Green, et al., 1994) which result from a defect in the control of cellular growth. If this were the case, however, we might expect that a proportion of constitutional PKD1 mutations would be inactivating deletions as seen in these other disorders.

The location of the PKD1 mutations may, however, reflect some ascertainment bias as it is this single copy area which has been screened most intensively for mutations. Nevertheless, no additional deletions were detected when a large part of the gene was screened by FIGE, and studies by PFGE showed no large deletions of this area in 75 PKD1 patients. It is possible that the mutations detected so far result in the production of an abnormal protein which causes disease through a gain of function. However, it is also possible that these mutations eliminate the production of functional protein from this chromosome and result in the PKD1

phenotype by haploinsufficiency, or only after loss of the second PKD1 homologue by somatic mutation.

At least one mutation which seems to delete the entire PKD1 gene has been identified (WS-53) but in this case it also disrupts the adjacent TSC2 gene and the resulting phenotype is of TSC with severe cystic kidney disease. Renal cysts are common in TSC so that the phenotypic significance of deletion of the PKD1 gene in this case is difficult to assess. It is clear that not all cases of renal cystic disease in TSC are due to disruption of the PKD1 gene; chromosome 9 linked TSC (TSC1) families also manifest cystic kidneys and we have analysed many TSC2 patients with kidney cysts who do not have deletion of the PKD1 gene.

Preliminary analysis of the PKD1 protein sequence has highlighted two regions which provide some clues to the possible function of the PKD1 gene. At the extreme 5' end of the characterised region are two leucine-rich repeats (LRRs) (amino acids 29-74) flanked by characteristic amino flanking (amino acids 6-28) and carboxy flanking sequences (amino acids 76-133) (Rothberg et al, 1990). LRRs are thought to be involved in protein-protein interations (Kobe and Deisenhofer, 1994) and the flanking sequences are only found in extracellular proteins. Other proteins with LRRs flanked on the amino and carboxy sides are receptors or are involved in adhesion or cellular signalling. Further 3' on the protein (amino acids 350-515) is a C-type lectin domain (Curtis et al, 1992). This indicates that this region binds carbohydrates and is also likely to be extracellular. These two regions of homology indicate that the 5' part of the PKD1 protein is extracellular and involved in protein-protein interactions. It is possible that this protein is a constituent of, or plays a role in assembling, the extracellular matrix (ECM) and may act

as an adhesive protein in the ECM. It is also possible that the extracellular portion of this protein is important in signalling to other cells. The function of much of the PKD1 protein is still not fully known but the presence of several hydrophobic regions indicates that the protein may be threaded through the cell membrane.

Familial studies indicate that de novo mutations probably account for only a small minority of all ADPKD cases; a recent study detected 5 possible new mutations in 209 families (Davies, et al., 1991). However, in our study one of three intragenic mutations detected was a new mutation and the PKD1 associated translocation was also a de novo event. Furthermore, the mutations detected in the two familial cases do not account for a significant proportion of the local PKD1. The OX875 deletion was only detected in 1 of 282 unrelated cases, and the splicing defect was seen in only 1 of 48 unrelated cases. Nevertheless, studies of linkage disequilibrium have found evidence of common haplotypes associated with PKD1 in a proportion of some populations (Peral, et al., 1994; Snarey, et al., 1994) suggesting that common mutations will be identified.

Once a larger range of mutations have been characterised it will be possible to evaluate whether the type and location of mutation determines disease severity, and if there is a correlation between mutation and extra-renal manifestations. Previous studies have provided some evidence that the risk of cerebral aneurysms 'runs true' in families (Huston, et al., 1993) and that some PKD1 families exhibit a consistently mild phenotype (Ryynanen, et al., 1987). A recent study has concluded that there is evidence of anticipation in ADPKD families, especially if the disease is transmitted through the mother (Fink, et

al., 1994). Furthermore, analysis of families with early manifestation of ADPKD show that there is a significant intra-familial recurrence risk and that childhood cases are most often transmitted maternally (Fink, et al., 1993; Zerres, et al., 1993). This pattern of inheritence is reminiscent of that seen in diseases in which an expanded trinucleotide repeat was found to be the mutational mechanism (reviewed in Mandel, 1993). However, no evidence for an expanding repeat correlating with PKD1 has been found in this region although such a sequence cannot be excluded.

There is ample evidence that early presymptomatic diagnosis of PKD1 is helpful because it allows complications such as hypertension and urinary tract infections to be monitored and treated quickly (Ravine, et al., 1991). The identification of mutations within a family will allow rapid screening of that and other families with the same mutation. However, genetic linkage analysis is likely to remain important for presymptomatic diagnosis. The accuracy and ease of linkage based diagnosis will be improved by the identification of the PKD1 gene as a microsatellite lies in the 3' untranslated region of this gene (KG-8) and several CA repeats are located 5' of the gene (see Figure 1a and 6; Peral, et al., 1994; Snarey, et al., 1994).

Experimental Procedures

Clinical Details of Patients

Family 77

77-2 and 77-3 are 48 and 17 years old, respectively, and have typical ADPKD. Both have bilateral polycystic kidneys and 77-2 has impaired renal function. Neither patient manifests any signs of TSC (apart from cystic kidneys) on clinical and ophthalmological examination or by CT scan of the brain.

77-4 is 13 years old, severely mentally retarded and has multiple signs of TSC including adenoma sebaceum, depigmented macules and periventricular calcification on CT scan. Renal ultrasound reveals a small number of bilateral renal cysts.

ADPKD patients

OX875 developed ESRD from ADPKD, aged 46. Progressive decline in renal function had been observed over 17 years; ultrasound examinations documented enlarging polycystic kidneys with less extensive hepatic cystic disease. Both kidneys were removed after renal transplantation and pathological examination showed typical advanced cystic disease in kidneys weighing 1920g and 3450g (normal average 120g).

OX114 developed ESRD from ADPKD aged 54: diagnosis was made by radiological investigation during an episode of abdominal pain aged 25. A progressive decline in renal function and the development of hypertension was subsequently observed. Ultrasonic examination demonstrated enlarged kidneys with typical cystic disease, with less severe hepatic involvement.

OX32 is a member of a large kindred affected by typical ADPKD in which several members have developed ESRD. The patient himself has been observed for 12 years with progressive renal failure and hypertension following ultrasonic demonstration of polycystic kidneys.

No signs of TSC were observed on clinical examination of any of the ADPKD patients.

DNA Electrophoresis and Hybridisation

DNA extraction, restriction digests, electrophoresis, Southern blotting, hybridisation and washing were performed by standard methods or as previously described (Harris, et al., 1990). FIGE was performed with the Biorad FIGE Mapper using programme 5 to separate fragments from 25-50 kb. High molecular

weight DNA for PFGE was isolated in agarose blocks and separated on the Biorad CHEF DRII apparatus using appropriate conditions.

Genomic DNA probes and somatic cell hybrids

Many of the DNA probes used in this study have been described previously: MS205.2 (D16S309; Royle, et al., 1992); GGG1 (D16S259; Germino, et al., 1990); N54 (D16S139; Himmelbauer, et al., 1991); SM6 (D16S665), CW23, CW21, and JH1 (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). Microsatellite probes for haplotype analysis were KG8 and W5.2 (Snarey, et al., 1994) SM6, CW3 and CW2, (Peral, et al., 1994), 16AC2.5 (Thompson, et al., 1992); SM7 (Harris, et al., 1991), VK5AC (Aksentijevich, et al., 1993).

New probes isolated during this study were: JH4, JH5, JH6, 11 kb, 6 kb and 6 kb BamH I fragments, respectively, and JH13 and JH14, 4 kb and 2.8 kb BamH I-EcoR I fragments, respectively, all from the cosmid JH2A; JH8 and JH10 are 4.5 kb and 2 kb Sac I fragments, respectively and JH12 a 0.6 Sac I-BamH I fragment, all from JH4; 8S1 and 8S3 are 2.4 kb and 0.6 kb Sac II fragments, respectively, from JH8; CW10 is a 0.5 kb Not I-Mlu I fragment of SM25A; JH17 is a 2 kb EcoR I fragment of NM17.

The somatic cell hybrids N-OH1 (Germino, et al., 1990), P-MWH2A (European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and Hy145.19 (Himmelbauer, et al., 1991) have previously been described. Somatic cell hybrids containing the paternally derived (BP2-10) and maternally derived (BP2-9) chromosomes from OX114 were produced by the method of Deisseroth and Hendrick (1979).

Constructing a cosmid contig

Cosmids were isolated from chromosome 16 specific and total genomic libraries, and a contig was constructed using the methods and libraries previously

described (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). To ensure that cosmids were derived from the 16p13.3 region (not the duplicate 16p13.1 area) initially, probes from the single copy area were used to screen libraries (e.g. CW21 and N54). Two cosmids mapped entirely within the area duplicated, CW10III and JC10.2B. To establish that these were from the PKD1 area, they were restriction mapped and hybridised with the probe CW10. The fragment sizes detected were compared to results obtained with hybrids containing only the 16p13.3 area (Hy145.19) or only the 16p13.1 region (P-MWH2A).

FISH

FISH was performed essentially as previously described (Buckle and Rack, 1993). The hybridisation mixture contained 100 ng of biotin-II-dUTP labelled cosmid DNA and 2.5 mg human Cot-1 DNA (BRL), which was denatured and annealled at 37°C for 15 min prior to hybridisation at 42°C overnight. After stringent washes the site of hybridisation was detected with successive layers of fluorescein-conjugated avidin (5 mg/ml) and biotinylated anti-avidin (5 mg/ml) (Vector Laboratories). Slides were mounted in Vectashield (Vector Laboratories) containing 1 mg/ml propidium iodide and 1 mg/ml 4', 6-diamidino-2-phenylindole (DAPI), to allow concurrent G-banded analysis under UV light. Results were analysed and images captured using a Bio-Rad MRC 600 confocal laser scanning microscope.

cDNA screening and characterisation

Foetal brain cDNAs libraries in λ phage (Clonetech and Stratagene) were screened by standard methods with genomic fragments in the single copy area (equivalent to CW23 and CW21) or with a 0.8 kb Pvu II-Eco RI single copy fragment of AH3. Six PBP cDNAs were characterised including two previously described, AH4 (1.7 kb) , 3A3 (2.0 kb) (European Chromosome 16 Tuberous Sclerosis

Consortium, 1993), and four novel cDNAs AH3 (2.2 kb), AH6 (2.0 kb), A1C (2.2 kb) and B1E (2.9 kb). A Striatum library (Stratagene) was screened with JH4 and a HG-C cDNA , 11BHS21 (3.8 kb) was isolated; 21P.9 is a 0.9 kb Pvu II-EcoR I subclone of this cDNA. A HG-A or HG-B cDNA, HG-4 (7 kb) was also isolated by screening the foetal brain library (Stratagene) with JH8. HG-4/1.1 is a 1.1 kb Pvu II-EcoR I fragment from the 3' end of HG-4. 1A1H.6 is a 0.6 kb Hind III-EcoR I subclone of a TSC2 cDNA, 1A-1 (1.7 kb), which was isolated from the Clonetech library. Each cDNA was subcloned into Bluescript and sequenced utilising a combination of sequential truncation and oligonucleotide primers using DyeDeoxy Terminators (Applied Biosystems) and an ABI 373A DNA Sequencer (Applied Biosystems) or by hand with 'Sequenase' T7 DNA polymerase (USB).

RNA Procedures

Total RNA was isolated from cell lines and tissues by the method of Chomczynski and Sacchi (1987) and enrichment for mRNA made using the PolyAT tract mRNA Isolation System (Promega). For RNA electrophoresis 0.5% agarose denaturing formaldehyde gels were used which were Northern blotted, hybridised and washed by standard procedures. The 0.24 - 9.5 kb RNA (Gibco BRL) size standard was used and hybridisation of the probe (1-9B3) to the 13 kb Utrophin transcript (Love, et al., 1989) in total fibroblast RNA was used as a size marker for the large transcripts.

RT-PCR was performed with 2.5 mg of total RNA by the method of Brown et al (1990) with random hexamer primers, except that AMV-reverse transcriptase (Life Sciences) was employed. To characterise the deletion of the PBP transcript in OX114 we used the primers :

AH3 F9 5' TTT GAC AAG CAC ATC TGG CTC TC 3'
AH3 B7 5' TAC ACC AGG AGG CTC CGC AG 3'

in a DMSO containing PCR buffer (Dodé, et al., 1990) with 0.5 mM $MgCl_2$ and 36 cycles of: 94°C, 1 min; 61°C, 1 min; 72°C, 2 min plus a final extension of 10 min. The 3A3 C primers used to amplify the OX32 cDNA and DNA were:

3A3 C1 5' CGC CGC TTC ACT AGC TTC GAC 3'
3A3 C2 5' ACG CTC CAG AGG GAG TCC AC 3'

These were employed in a PCR buffer and cycle previously described (Harris, et al., 1991) with 1mM $MgCl_2$ and an annealing temperature of 61°C.

PCR products for sequencing were amplified with Pfu-1 (Stratagene) and ligated into the Srf-1 site in PCR-Script (Stratagene) in the presence of Srf-1.

RNAse protection

Tissues from normal and end-stage polycystic kidneys were immediately homogenised in guanidinium thiocyanate. RNA was purified on a cesium chloride gradient and 30 mg total RNA was assayed by RNAse protection by the method of Melton, et al., (1984) using a genomic template generated with the 3A3, C primers.

Heteroduplex Analysis

Heteroduplex analysis was performed essentially as described by Keen et al (1991). Samples were amplified from genomic DNA with the 3A3, C primers, heated at 95°C for 5 minutes and incubated at room temperature for at least 30 minutes before loading on a Hydrolink gel (AT Biochem). Hydrolink gels were run for 12-18 hours at 250V and fragments observed after staining with ethidium bromide.

Extraction and amplification of paraffin-embedded DNA

DNA from formalin fixed, paraffin wax embedded kidney tissue was prepared by the method of Wright and

Manos (1990), except that after proteinase K digestion overnight at 55°C, the DNA was extracted with phenol plus chloroform before ethanol precipitation. Approximately 50 ng of DNA was used for PCR with 1.5 mM $MgCl_2$ and 40 cycles of 94°C for 1 min, 59°C for 1 min and 72°C for 40 s, plus a 10 min extension at 72°C. The oligonucleotide primers designed to amplify across the genomic deletion of OX875 were:

AH4F2 : 5' - GGG CAA GGG AGG ATG ACA AG - 3'

JH14B3 : 5' - GGG TTT ATC AGC AGC AAG CGG - 3'

which produced a product of ~ 220 bp in individuals with the OX875 deletion.

3'RACE analysis of WS-212

3' RACE was completed essentially as described (European Polycystic Kidney Disease Consortium (1994)). Reverse transcription was performed with 5μg total RNA with 0.5μg of the hybrid dT_{17} adapter primer using conditions previously described (Fronman et al., (1988)). A specific 3' RACE product was amplified with the primer F5 adn adapter primer in 0.5mM $MgCl_2$ with the program: 57°C, 60s; 72°C, 15 minutes and 30 cycles of 95°C, 40s; 57°C, 60s; 72°C, 60s plus 72°C, 10 minutes. The amplified product was cloned using the TA cloning system (Invitrogen) and sequenced by conventional methods.

References

Aksentijevich, I., Pras, E., Gruberg, L., Shen, Y., Holman, K., Helling, S., Prosen, L., Sutherland, G.R., Richards, R.I., Ramsburg, M., Dean, M., Pras, M., Amos, C.I. and Kastner, D.L. (1993). Refined mapping of the gene causing familial Mediterranean fever, by linkage and homozygosity studies. Am. J. Hum. Genet. 53, 451-461.

Altschul, S.F., Warren, G., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic alignment search tool. J. Mol. Biol. 215, 403-410.

Breuning, M.H., Reeders, S.T., Brunner, H., Ijdo, J.W., Saris, J.J., Verwest, A., van Ommen, G.J.B. and Pearson, P.L. (1987). Improved early diagnosis of adult polycystic kidney disease with flanking DNA markers. Lancet. ii, 1359-1361.

Breuning, M.H., Snijdewint, F.G.M., Brunner, H., Verwest, A., Ijdo, J.W., Saris, J.J., Dauwerse, J.G., Blonden, L., Keith, T., Callen, D.F., Hyland, V.J., Xiao, G.H., Scherer, G., Higgs, D.R., Harris, P., Bachner, L., Reeders, S.T., Germino, G., Pearson, P.L. and van Ommen, G.J.B. (1990). Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). J. Med. Genet. 27, 603-613.

Brook-Carter, P.T., Peral, B., Ward, C.J., Thompson, P., Hughes, J., Maheshwar, M.M., Nellist, M., Gamble, V., Harris, P.C. & Sampson, J.R. (1994). Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease - a contiguous gene syndrome. Nature Genetics, 8, 328-332.

Brown, C.J., Flenniken, A.M., Williams, B.R.G. and Willard, H.F. (1990). X chromosome inactivation of the human TIMP gene. Nucl. Acids Res. 18, 4191-4195.

Buckle, V.J. and Rack, K. (1993). Fluorescent in situ hybridisation. Human Genetic Disease Analysis: A Practical Approach (K.E. Davies, Ed.) Vol 2, 59-82. IRL Press, Oxford.

Chapman, A.B., Rubinstein, D., Hughes, R., Stears, J.C., Earnest, M.P., Johnson, A.M., Gabow, P.A. and Kaehny, W.D. (1992). Intracranial aneurysms in autosomal dominant polycystic kidney disease. N. Eng. J. Med. 327, 916-920.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156-159.

Curtis, B.M., Scharnowske, S. & Watson, A.J., (1992) Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proceedings of the National Academy of Sciences, USA, 89, 8356-8360.

Dalgaard, O.Z. (1957). Bilateral polycystic disease of the kidneys: A follow-up of two hundred and eighty-four patients and their families. Acta Med. Scand. 158, 1-251.

Davies, F., Coles, G.A., Harper, P.S., Williams, A.J., Evans, C. and Cochlin, D. (1991). Polycystic kidney disease re-evaluated: a population-based study. Q. J. Med. 79, 477-485.

Deisseroth, A. and Hendrick, D. (1979). Activation of phenotypic expression of human globin genes from non-erythroid cells by chromosome-dependent transfer to tetraploid mouse erythroleukemia cells. Proc. Natl. Acad. Sci. USA. 76, 2185-2189.

Dodé, C., Rochette, J. and Krishnamoorthy, R. (1990). Locus assignment of human a globin mutations by selective amplification and direct sequencing. Brit. J. Haemat. 76, 275-281.

European Chromosome 16 Tuberous Sclerosis Consortium (1993). Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell. 75, 1305-1315.

European Polycystic Kidney Disease Consortium (1994) The polycystic kidney disease 1 gene encodes at 14 kb transcript and lies within a duplicated region on chromosome 16. Cell. 77, 881-894.

Fink, G.M., Johnson, A.M., Strain, J.D., Kimberling, W. J., Kumar, S., Manco-Johnson, M.L., Duley, I.T. and Gabow, P.A. (1993). Characteristics of very early onset autosomal dominant polycystic kidney disease. Journal American Society Nephrology. 3, 1863-1870.

Fink, G.M., Johnson, A.M. and Gabow, P.A. (1994). Is there evidence for anticipation in autosomal-dominant polycystic disease? Kidney Int. 45, 1153-1162.

Fronman, M.A., Dush, M.K. & Martin, G.R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligo-nucleotide primer. Biochemistry 85, 8998-9002 (1988).

Gabow, P.A. (1990). Autosomal dominant polycystic kidney disease - more than a renal disease. American Journal of Kidney Diseases. XVI, 403-413.

Gabow, P.A. (1991). Polycystic kidney disease: clues to pathogenesis. Kidney Int. 40, 989-996.

Gabow, P.A., Johnson, A.M., Kaehny, W.D., Kimberling, W.J., Lezotte, D.C., Duley, I.T. and Jones, R.H. (1992). Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease. Kidney Int. 41, 1311-1319.

Gabow, P.A. (1993). Autosomal dominant polycystic kidney disease. N. Eng. J. Med. 329, 332-342.

Germino, G.G., Barton, N.J., Lamb, J., Higgs, D.R., Harris, P., Xiao, G.H., Scherer, G., Nakamura, Y. and Reeders, S.T. (1990). Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. Am. J. Hum. Genet. 46, 925-933.

Germino, G.G., Weinstat-Saslow, D., Himmelbauer, H., Gillespie, G.A.J., Somlo, S., Wirth, B., Barton, N., Harris, K.L., Frischauf, A.-M. and Reeders, S.T. (1992). The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics. 13, 144-151.

Green, A.J., Smith, M. and Yates, J.R.W. (1994). Loss of heterozygosity on chromosome 16p13.3 in hamartomas from tuberous sclerosis patients. Nature Genet. 6, 193-196.

Harris, P.C., Barton, N.J., Higgs, D.R., Reeders, S.T. and Wilkie, A.O.M. (1990). A long-range restriction map between the a-globin complex and a marker closely linked to the polycystic kidney disease (PKD1) locus. Genomics. 7, 195-200.

Harris, P.C., Thomas, S., Ratcliffe, P.J., Breuning, M.H., Coto, E. and Lopez-Larrea, C. (1991). Rapid genetic analysis of families with polycystic kidney disease by means of a microsatellite marker. Lancet. 338, 1484-1487.

Himmelbauer, H., Germino, G.G., Ceccherini, I., Romeo, G., Reeders, S.T. and Frischauf, A.-M. (1991). Saturating the region of the polycystic kidney disease gene with Not I linking clones. Am. J. Hum. Genet. 48, 325-334.

Hossack, K.F., Leddy, C.L., Johnson, A.M., Schrier, R.W. and Gabow, P.A. (1988). Echocardiographic findings in autosomal dominant polycystic kidney disease. N. Eng. J. Med. 319, 907-912.

Huston, J., Torres, V.E., Sulivan, P.P., Offord, K.P. and Wiebers, D.O. (1993). Value of magnetic resonance angiography for detection of intracranial aneurysm in autosomal dominant polycystic kidney disease. Journal of the American Society of Nephrology. 3, 1871-1877.

Hyland, V.J., Suthers, G.K., Friend, K., MacKinnon, R.N., Callen, D.F., Breuning, M.H., Keith, T., Brown, V.A., Phipps, P. and Sutherland, G.R. (1990). Probe, VK5B, is located in the same interval as the autosomal dominant adult polycystic kidney disease locus, PKD1. Hum. Genet. 84, 286-288.

Keen, J., Lester, D., Inglehearn, C., Curtis, A. and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. Trends Genet. 7, 5.

Kimberling, W.J., Fain, P.R., Kenyon, J.B., Goldgar, D., Sujansky, E. and Gabow, P.A. (1988). Linkage heterogeneity of autosomal dominant polycystic kidney disease. N. Eng. J. Med. 319, 913-918.

Kimberling, W.J., Kumar, S., Gabow, P.A., Kenyon, J.B., Connolly, C.J. and Somlo, S. (1993). Autosomal dominant polycystic kidney disease: Localization of the second gene to chromosome 4q13-q23. Genomics. 18, 467-472.

Kobe, B. & Deisenhofer, J. (1994) The leucine-rich repeat: a versatile binding motif. Trends in Biochemical Sciences, **19**, 415-421.

Legius, E., Marchuk, D.A., Collins, F.S. and Glover, T.W. (1993). Somatic deletion of the neurofibromatosis type 1 gene in a neurofibrosarcoma supports a tumour suppressor gene hypothesis. Nature Genet. 3, 122-126.

Love, D.R., Hill, D.F., Dickson, G., Spurr, N.K., Byth, B.C., Marsden, R.F., Walsh, F.S., Edwards, Y.H. and Davies, K.E. (1989). An autosomal transcript in skeletal muscle with homology to dystrophin. Nature. 339, 55-58.

Mandel, J-L. Questions of expansion. Nature Genetics. 4, 8-9.

Melton, D.A., Kreig, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Efficient in vitro synthesis of biological active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acid Research. 12, 7035-7056.

Milutinovic, J., Fialkow, P.J., Rudd, T.G. and Agodoa, L.Y., Phillips, L.A. and Bryant, J.I. (1980). Liver cysts in patients with autosomal dominant polycystic kidney disease. Am. J. Med. 68, 741-744.

Parfrey, P.S., Bear, J.C., Morgan, J., Cramer, B.C., McManamon, P.J., Gault, M.H., Churchill, D.N., Singh, M., Hewitt, R., Somlo, S. and Reeders, S.T. (1990). The diagnosis and prognosis of autosomal dominant polycystic kidney disease. N. Eng. J. Med. 323, 1085-1090.

Peral, B., Ward, C.J., San Millán, J.L., Thomas, S., Stallings, R.L., Moreno, F. and Harris, P.C. (1994). Evidence of linkage disequilibrium in the Spanish polycystic kidney disease 1 population. Am. J. Hum. Genet. 54, 899-908.

Peters, D.J.M. and Sandkuijl, L.A. (1992). Genetic heterogeneity of polycystic kidney disease in Europe. Contributions to Nephrology: Polycystic Kidney Disease. 97, 128-139.

Peters, D.J.M., Spruit, L., Saris, J.J., Ravine, D., Sandkuijl, L.A., Fossdal, R., Boersma, J., van Eijk, R., N rby, S., Constantinou-Deltas, C.D., Pierides, A., Brissenden, J.E., Frants, R.R., van Ommen, G.-J.B. and Breuning, M.H. (1993). Chromosome 4 localization of a second gene for autosomal dominant polycystic kidney disease. Nature Genet. 5, 359-362.

Pound, S.E., Carothers, A.D., Pignatelli, P.M., Macnicol, A.M., Watson, M.L. and Wright, A.F. (1992). Evidence for linkage disequilibrium between D16S94 and the adult onset polycystic kidney disease (PKD1) gene. J. Med. Genet. 29, 247-248.

Ravine, D., Walker, R.G., Gibson, R.N., Forrest, S.M., Richards, R.I., Friend, K., Sheffield, L.J., Kincaid-Smith, P. and Danks, D.M. (1992). Phenotype and genotype heterogeneity in autosomal dominant polycystic kidney disease. Lancet. 340, 1330-1333.

Ravine, D., Walker, R.G., Gibson, R.N., Sheffield, L.J., Kincaid-Smith, P. and Danks, D.M. (1991). Treatable complications in undiagnosed cases of autosomal dominant polycystic kidney disease. Lancet. 337, 127-129.

Reeders, S.T. (1992). Multilocus polycystic disease. Nature Genet. 1, 235-237.

Reeders, S.T., Breuning, M.H., Davies, K.E., Nicholls, R.D., Jarman, A.P., Higgs, D.R., Pearson, P.L. and Weatherall, D.J. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature. 317, 542-544.

Reeders, S.T., Keith, T., Green, P., Germino, G.G., Barton, N.J., Lehmann, O.J., Brown, V.A., Phipps, P., Morgan, J., Bear, J.C. and Parfrey, P. (1988). Regional localization of the autosomal dominant polycystic kidney disease locus. Genomics. 3, 150-155.

Romeo, G., Costa, G., Catizone, L., Germino, G.G., Weatherall, D.J., Devoto, M., Roncuzzi, L., Zucchelli, P., Keith, T. and Reeders, S.T. (1988). A second genetic locus for autosomal dominant polycystic kidney disease. Lancet. ii, 8-10.

Rothberg, J.M., Jacobs, J.R., Goodman, C.S. & Artavanis-Tsakonas, S. (1990) slit: an extracellular protein necessary for development of midlie glia and

commissural axon pathways contains both EGF and LRR domains. Genes and Development, 4, 2169-2187.

Royle, N.J., Armour, J.A.L., Webb, M., Thomas, A. and Jeffreys, A.J. (1992). A hypervariable locus D16S309 located at the distal end of 16p. Nucl. Acids Res. 20, 1164.

Ryynanen, M., Dolata, M.M., Lampainen, E. and Reeders, S.T. (1987). Localisation of a mutation producing autosomal dominant polycystic kidney disease without renal failure. J. Med. Genet. 24, 462-465.

Scheff, R.T., Zuckerman, G., Harter, H. Delmez, J. and Koehler, R.(1980). Diverticular disease in patients with chronic renal failure due to polycystic kidney disease. Ann. Intern. Med. 92, 202-204.

Snarey, A., Thomas, S., Schneider, M.C., Pound, S.E., Barton, N., Wright, A.F., Harris, P.C., Reeders, S.T. and Frischauf, A.-M. (1994). Linkage disequilibrium in the region of the autosomal dominant polycystic kidney disease gene (PKD1). Am. J. Hum. Genet. in press.

Somlo, S., Wirth, B., Germino, G.G., Weinstat-Saslow, D., Gillespie, G.A.J., Himmelbauer, H., Steevens, L., Coucke, P., Willems, P., Bachner, L., Coto, E., Lopez-Larrea, C., Peral, B., San Millan, J.L., Saris, J.J., Breuning, M.H., Frischauf, A.-M. and Reeders, S.T. (1992). Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (PKD1) with respect to physically mapped markers. Genomics. 13, 152-158.

Thompson, A.D., Shen, Y., Holman, K., Sutherland, G.R., Callen, D.F. and Richards, R.I. (1992). Isolation and characterization of $(AC)_n$ microsatellite genetic markers from human chromosome 16. Genomics. 13, 402-408.

Wilson, P.D. and Sherwood, A.C. (1991). Tubulocystic epithelium. Kidney Int. 39, 450-463.

Wright, D.K. and Manos, M.M. (1990). Sample preparation from paraffin-embedded tissues. PCR Protocols: A Guide to Methods and Applications. 153-166.

Zerres, K., Rudnik-Schöneborn, S., Deget, F. and German working group on paediatric nephrology (1993). Childhood onset autosomal dominant polycystic kidney disease in sibs: clinical picture and recurrence risk. J. Med. Genet. 30, 583-588.

CLAIMS

1. An isolated, purified or recombinant nucleic acid sequence comprising:-

(a) a PKD1 gene or its complementary strand,

(b) a sequence substantially homologous to, or capable of hybridising to, a substantial portion of a molecule defined in (a) above,

(c) a fragment of a molecule defined in (a) or (b) above.

2. A sequence according to claim 1, wherein the PKD1 gene has the partial nucleic acid sequence according to Figure 7 and/or 10.

3. A sequence according to claim 1 or claim 2 comprising a DNA molecule selected from:

(a) a PKD1 gene or its complementary strand,

(b) a sequence substantially homologous to, or capable of hybridising to, a substantial portion of a molecule defined in (a) above,

(c) a molecule coding for a polypeptide having the partial sequence of Figure 7,

(d) genomic DNA corresponding to a molecule in (a) above; and

(e) a fragment of a molecule defined in any of (a), (b), (c) or (d) above.

4. A nucleic acid sequence comprising a mutant PKD1 gene, selected from those wherein:-

(a) [OX114] base pairs 1746-2192 as defined in

Figure 7 are deleted (446bp);

(b) [OX32] base pairs 3696-3831 as defined in Figure 7 are deleted by a splicing defect;

(c) [OX875] about 5.5kb flanked by the two XbaI sites shown in Figure 3a are deleted and the EcoR1 site separating the CW10 (41kb) and JH1 (18kb) sites is thereby absent; and

(d) [WS53] about 100kb extending between the JH1 and CW21 and the SM6 and JH17 sites shown in Figure 6 and the PKD1 gene is thereby absent.

5. A nucleic acid sequence comprising a mutant PKD1 gene selected from those wherein-

(a) [461] abpout 18bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(b) [OX1054] about 20bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(c) [WS212] about 75kb are deleted between SM9-CW9 distally and the PKD1 3'UTR proximally as shown in Figure 12;

(d) [WS-215] about 160kb are deleted between CW20 and CW10-CW36 as shown in Figure 12;

(e) [WS-227] about 50kb are deleted between CW20 and JH11 as shown in Figure 12;

(f) [WS-219] about 27kb are deleted between JH1 and JH6 as shown in Figure 12; and

(g) [WS-250] about 160kb are deleted betwenn WC20

and BLu24 as shown in Figure 12.

(h) [WS194] a deletion of about 65kb between CW20 and CW10.

6. An RNA molecule comprising an RNA sequence corresponding to a DNA sequence according to any of claims 1 to 5.

7. An RNA molecule according to claim 6, wherein the molecule is the transcript referenced PKD1 and identifiable from the restriction map of Figure 3a and having a sequence of about 14 KB.

8. A nucleic acid probe having a sequence according to any of the preceding claims and optionally including a label.

9. A nucleic acid sequence according to any preceding claim, wherein the nucleic acid sequence encoding PKD1 is operably linked to transcriptional and/or translational expression signals.

10. An isolated, purified or recombinant polypeptide comprising a PKD1 protein or a mutant or variant thereof or encoded by a sequence according to any of claims 1 to 9 or a variant thereof having substantially the same activity as the PKD1 protein.

11. A polypeptide according to claim 10, wherein the PKD1 protein has the amino acid sequence according to the partial amino acid sequence of Figure 7 and/or Figure 10.

12. An anti-PKD1 antibody or a labelled anti-PKD1 antibody.

13. A method for screening a subject to determine

whether said subject is a PKD1-associated disorder carrier or a patient having a PKD1-associated disorder, which method comprises detecting the presence of and/or evaluating the characteristics of PKD1 DNA, PKD1 RNA and/or PKD1 polypeptide in a biological sample from said patient.

14. A method according to claim 13 which is or includes detecting and/or evaluating whether the PKD1 DNA is mutated, deleted, aberrant or otherwise abnormal, or is not expressing normal PKD1 protein.

15. A method according to claim 13 or claim 14, wherein the detection and/or evaluation includes the step of comparing the results thereof with results obtained using a mutated sequence according to claim 4 or claim 5.

16. A method according to any of claims 13 to 15, wherein said screening includes applying a nucleic acid amplification process to said sample to amplify a fragment of the PKD1 DNA or cDNA corresponding to the PKD1 RNA.

17. A method according to claim 16, wherein said nucleic acid amplification process uses at least one of the following sets of primers as identified herein:-

AH3 F9 : AH3 B7

3A3 C1 : 3A3 C2

AH4 F2 : JH14 B3

18. A method according to any of claims 13 to 17 which comprises digesting said sample to EcoR1 fragments and hybridising with a DNA probe which hybridises to the EcoR1 fragment identified (A) in Figure 3(a).

19. A method according to claim 18, wherein said DNA

probe comprises the DNA probe CW10 identified herein.

20. A method according to any of claims 13 to 17 which comprises digesting said sample to provide BamHl fragments hybridising with a DNA probe which hybridises to the BamHl fragment identified (B) in Figure 3(a).

21. A method according to claim 20, wherein said DNA probe comprises the DNA probe 1A1H.6 identified herein.

22. A vector (such as Bluscript (available from Stratagene)) comprising the nucleic acid sequence of any of claims 1 to 9.

23. A host cell (such as E. coli strain SL-1 Blue (available from Stratgene)) transfected or transformed with a vector according to claim 22.

24. The use of a vector according to claim 23 or a nucleic acid sequence according to any of claims 1 to 11 in gene therapy and/or in the preparation of an agent for treating or preventing a PKD1-associated disorder.

25. A method of treating or preventing a PKD1-associated disorder which method comprises administering to a patient in need thereof a functional PKD1 gene to affected cells in a manner that permits expression of PKD1 protein therein and/or a transcript produced from a mutated chromosome such as the deleted WS-212 chromosome which is capable of expressing functional PKD1 protein therein.

26. A diagnostic kit for carrying out a method according to any of claims 13 to 21, comprising nucleic acid primers for amplifying a fragment of a sequence according to any of Claims 1 to 9.

27. A diagnostic kit according to claim 26, wherein the nucleic acid primers comprise at least one of the following sets:

AH3 F9 : AH3 B7

3A3 C1 : 3A3 C2

AH4 F2 : JH14 B3

28. A diagnostic kit for carrying out a method according to claim 18, including one or more substances for digesting a sample to provide EcoRI fragments and a DNA probe as defined in claim 19.

29. A diagnostic kit for carrying out a method according to claim 20, including one or more substances for digesting a sample to provide BamHl fragments and a DNA probe as defined in claim 21.

30. A diagnostic kit for carrying out a method for determining whether said subject is a PKD1-associated disorder carrier or a patient having a PKD1-associated disorder, which includes a nucleic acid probe capable of hybridising to a sequence according to any of claims 1 to 11.

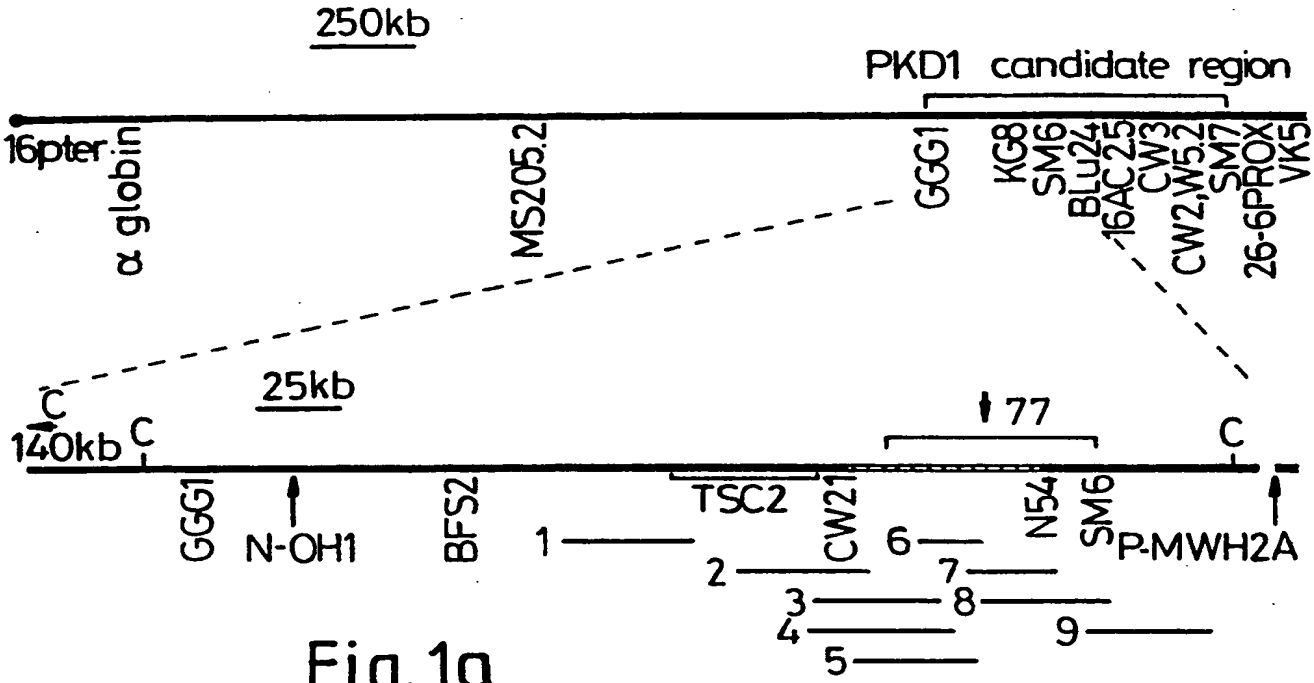

Fig. 1a

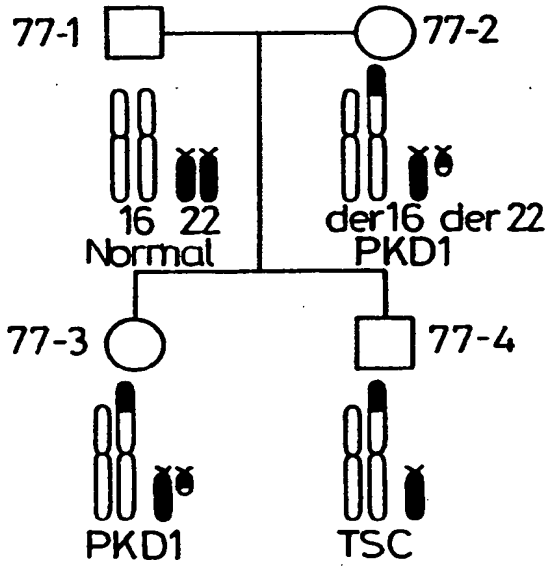

Fig. 1b

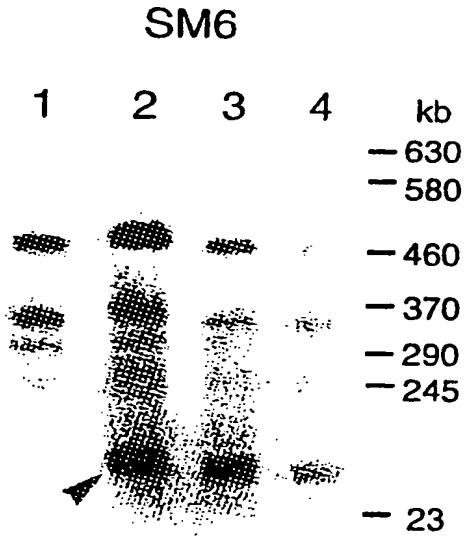

Fig. 1c

Fig. 2

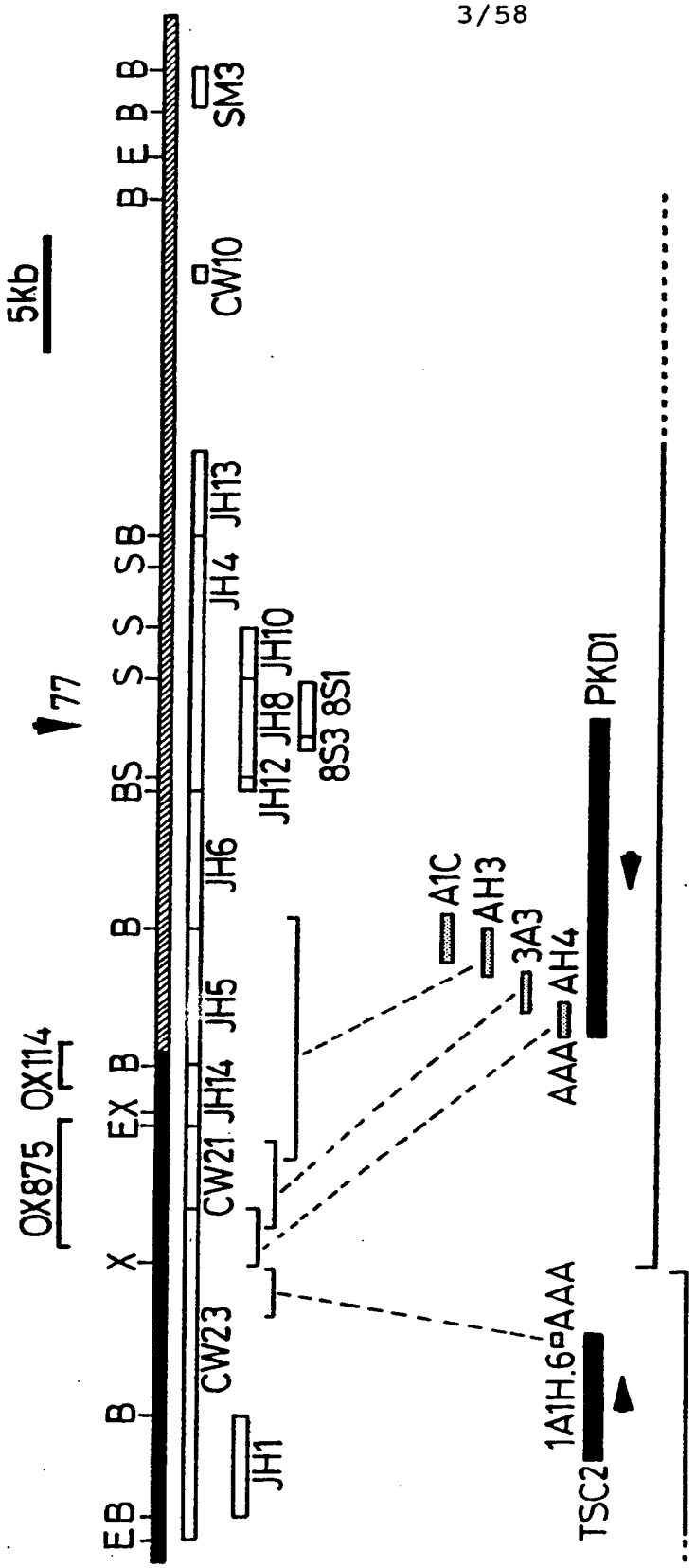

Fig. 3a

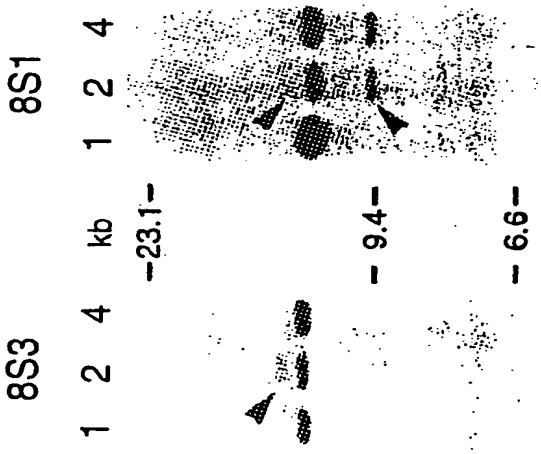

Fig. 3b

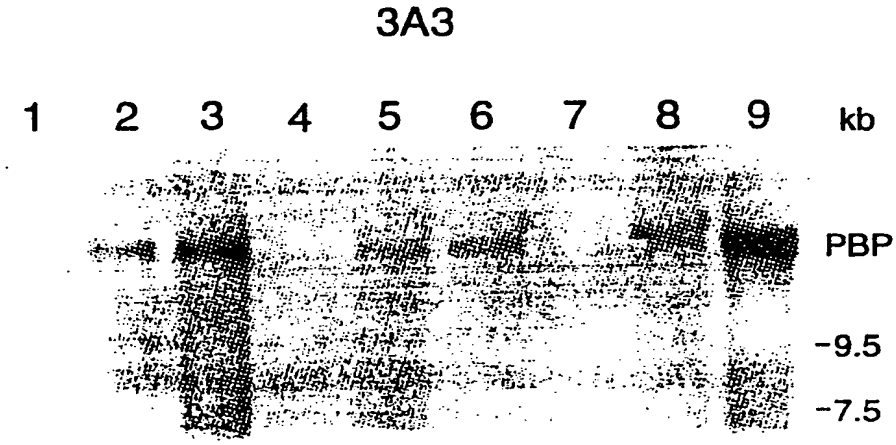

Fig. 4a

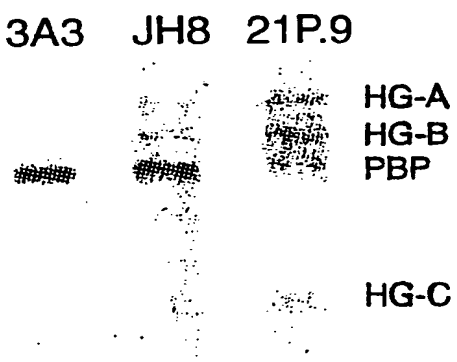

Fig. 4b

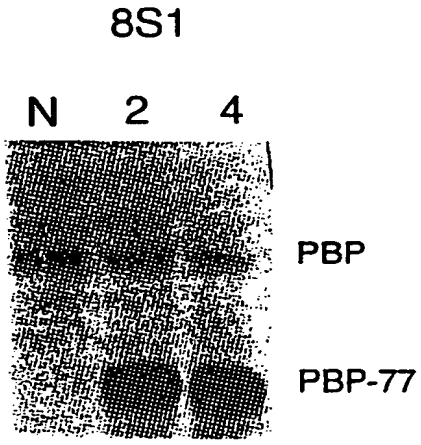

Fig. 4c

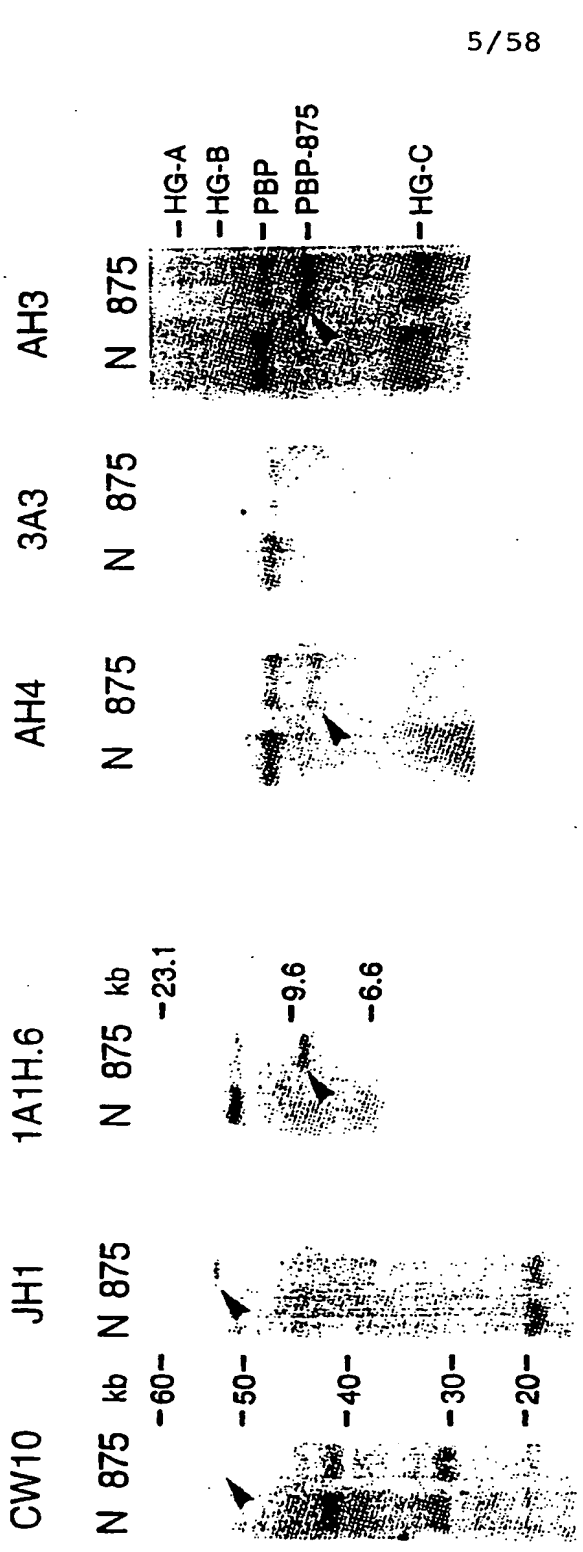

Fig. 5a

Fig. 5b

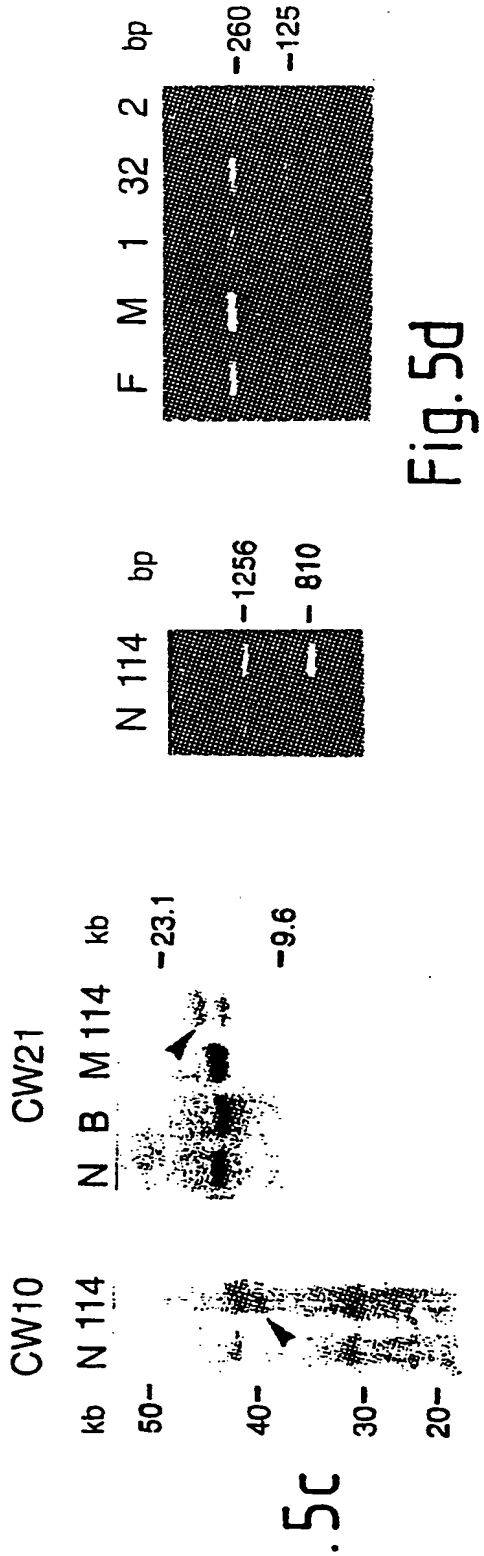

Fig. 5c

Fig. 5d

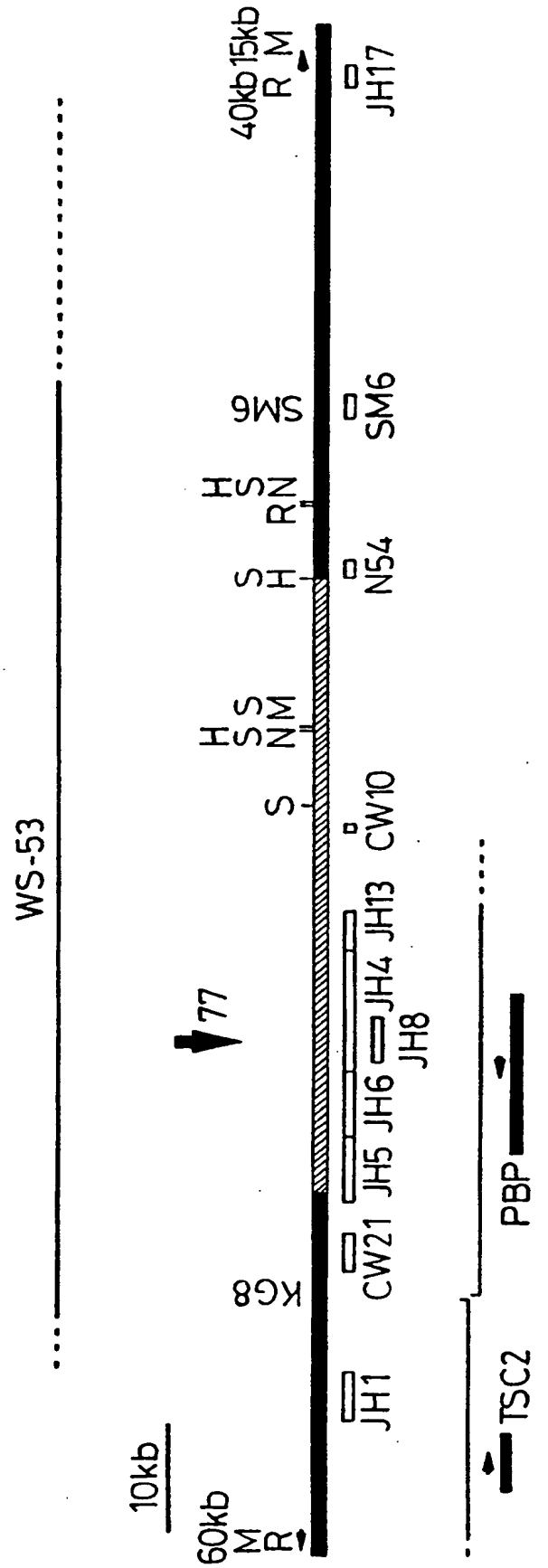

Fig. 6

```
   1  CTCAACGAGGAGCCCCTGACGCTGGCGGGCGAGGAGATCGTGGCCCAGGGCAAGCGCTCG   60
   1  L  N  E  E  P  L  T  L  A  G  E  E  I  V  A  Q  G  K  R  S   20

  61  GACCCGCGGAGCCTGCTGTGCTATGGCGGCGCCCCAGGGCCTGGCTGCCACTTCTCCATC  120
  21  D  P  R  S  L  L  C  Y  G  G  A  P  G  P  G  C  H  F  S  I   40

 121  CCCGAGGCTTTCAGCGGGGCCCTGGCCAACCTCAGTGACGTGGTGCAGCTCATCTTTCTG  180
  41  P  E  A  F  S  G  A  L  A  N  L  S  D  V  V  Q  L  I  F  L   60

 181  GTGGACTCCAATCCCTTTCCCTTTGGCTATATCAGCAACTACACCGTCTCCACCAAGGTG  240
  61  V  D  S  N  P  F  P  F  G  Y  I  S  N  Y  T  V  S  T  K  V   80

 241  GCCTCGATGGCATTCCAGACACAGGCCGCCGCCCAGATCCCCATCGAGCGGCTGGCCTCA  300
  81  A  S  M  A  F  Q  T  Q  A  G  A  Q  I  P  I  E  R  L  A  S  100

 301  GAGCGCGCCATCACCGTGAAGGTGCCCAACAACTCGGACTGGGCTGCCCGGGGCCACCGC  360
 101  E  R  A  I  T  V  K  V  P  N  N  S  D  W  A  A  R  G  H  R  120

 361  AGCTCCGCCAACTCCGCCAACTCCGTTGTGGTCCAGCCCCAGGCCTCCGTCGGTGCTGTG  420
 121  S  S  A  N  S  A  N  S  V  V  V  Q  P  Q  A  S  V  G  A  V  140

 421  GTCACCCTGGACAGCAGCAACCCTGCGGCCGGGCTGCATCTGCAGCTCAACTATACGCTG  480
 141  V  T  L  D  S  S  N  P  A  A  G  L  H  L  Q  L  N  Y  T  L  160

 481  CTGGACGGCCACTACCTGTCTGAGGAACCTGAGCCCTACCTGGCAGTCTACCTACACTCG  540
 161  L  D  G  H  Y  L  S  E  E  P  E  P  Y  L  A  V  Y  L  H  S  180

 541  GAGCCCCGGCCCAATGAGCACAACTGCTCGGCTAGCAGGAGGATCCGCCCAGAGTCACTC  600
 181  E  P  R  P  N  E  H  N  C  S  A  S  R  R  I  R  P  E  S  L  200

 601  CAGGGTGCTGACCACCGGCCCTACACCTTCTTCATTTCCCCGGGGAGCAGAGACCCAGCG  660
 201  Q  G  A  D  H  R  P  Y  T  F  F  I  S  P  G  S  R  D  P  A  220

 661  GGGAGTTACCATCTGAACCTCTCCAGCCACTTCCGCTGGTCGGCGCTGCAGGTGTCCGTG  720
 221  G  S  Y  H  L  N  L  S  S  H  F  R  W  S  A  L  Q  V  S  W  240

 721  GGCCTGTACACGTCCCTGTGCCAGTACTTCAGCGAGGAGGACATGGTGTGGCGGACAGAG  780
 241  G  L  Y  T  S  L  C  Q  Y  F  S  E  E  D  M  V  W  R  T  E  260

 781  GGGCTGCTGCCCCTGGAGGAGACCTCGCCCCGCCAGGCGGTCTGCCTCACCGGCACCTC  840
 261  G  L  L  P  L  E  E  T  S  P  R  Q  A  V  C  L  T  R  H  L  280

 841  ACCGCCTTCGGCGCCAGCCTCTTCGTGCCCCCAAGCCATGTCCGCTTTGTGTTTCCTGAG  900
 281  T  A  F  G  A  S  L  F  V  P  P  S  H  V  R  F  V  F  P  E  300

 901  CCGACAGCGGATGTAAACTACATCGTCATGCTGACATGTGCTGTGTGCCTGGTGACCTAC  960
 301  P  T  A  D  V  N  Y  I  V  M  L  T  C  A  V  C  L  V  T  Y  320

 961  ATGGTCATGGCCGCCATCCTGCACAAGCTGGACCAGTTGGATGCCAGCCGGGGCCGCGCC 1020
 321  M  V  M  A  A  I  L  H  K  L  D  Q  L  D  A  S  R  G  R  A  340

1021  ATCCCTTTCTGTGGGCAGCGGGGCCGCTTCAAGTACGAGATCCTCGTCAAGACAGGCTGG 1080
 341  I  P  F  C  G  Q  R  G  R  F  K  Y  E  I  L  V  K  T  G  W  360

1081  GGCCGGGGCTCAGGTACCACGGCCCACGTGGGCATCATGCTGTATGGGGTGGACAGCCGG 1140
 361  G  R  G  S  G  T  T  A  H  V  G  I  M  L  Y  G  V  D  S  R  380

1141  AGCGGCCACCGGCACCTGGACGGCGACAGAGCCTTCCACCGCAACAGCCTGGACATCTTC 1200
 381  S  G  H  R  H  L  D  G  D  R  A  F  H  R  N  S  L  D  I  F  400

1201  CGGATCGCCACCCCGCACAGCCTGGGTAGCGTGTGGAAGATCCGAGTGTGGCACGACAAC 1260
 401  R  I  A  T  P  H  S  L  G  S  V  W  K  I  R  V  W  H  D  N  420
```

Figure 7

```
1261  AAAGGGCTCAGCCCTGCCTGGTTCCTGCAGCACGTCATCGTCAGGGACCTGCAGACGGCA  1320
 421  K  G  L  S  P  A  W  F  L  Q  H  V  I  V  R  D  L  Q  T  A   440

1321  CGCAGCGCCTTCTTCCTGGTCAATGACTGGCTTTCGGTGGAGACGGAGGCCAACGGGGGC  1380
 441  R  S  A  F  F  L  V  N  D  W  L  S  V  E  T  E  A  N  G  G   460

1381  CTGGTGGAGAAGGAGGTGCTGGCCGCGAGCGACGCAGCCCTTTTGCGCTTCCGGCGCCTG   140
 461  L  V  E  K  E  V  L  A  A  S  D  A  A  L  L  R  F  R  R  L   480

1441  CTGGTGGCTGAGCTGCAGCGTGGCTTCTTTGACAAGCACATCTGGCTCTCCATATGGGAC  1500
 481  L  V  A  E  L  Q  R  G  F  F  D  K  H  I  W  L  S  I  W  D   500

1501  CGGCCGCCTCGTAGCCGTTTCACTCGCATCCAGAGGGCCACCTGCTGCGTTCTCCTCATC  1560
 501  R  P  P  R  S  R  F  T  R  I  Q  R  A  T  C  C  V  L  L  I   520

1561  TGCCTCTTCCTGGGCGCCAACGCCGTGTGGTACGGGGCTGTTGGCGACTCTGCCTACAGC  1620
 521  C  L  F  L  G  A  N  A  V  W  Y  G  A  V  G  D  S  A  Y  S   540

1621  ACGGGGCATGTGTCCAGGCTGAGCCCGCTGAGCGTCGACACAGTGGCTGTTGGCCTCGTG  1680
 541  T  G  H  V  S  R  L  S  P  L  S  V  D  T  V  A  V  G  L  V   560

1681  TCCAGCGTGGTTGTCTATCCCGTCTACCTGGCCATCCTTTTTCTCTTCCGGATGTCCCGG  1740
 561  S  S  V  V  V  Y  P  V  Y  L  A  I  L  F  L  F  R  M  S  R   580

1741  AGCAAGGTGGCTGGGAGCCCGAGCCCCCACACTGCCGGGCAGCAGGTGCTGGACATCGAC  1800
 581  S  K  V  A  G  S  P  S  P  T  P  A  G  Q  Q  V  L  D  I  D   600

1801  AGCTGCCTGGACTCGTCCGTGCTGGACAGCTCCTTCCTCACGTTCTCAGGCCTCCACGCT  1860
 601  S  C  L  D  S  S  V  L  D  S  S  F  L  T  F  S  G  L  H  A   620

1861  GAGGCCTTTGTTGGACAGATGAAGAGTGACTTGTTTCTGGATGATTCTAAGAGTCTGGTG  1920
 621  E  A  F  V  G  Q  M  K  S  D  L  F  L  D  D  S  K  S  L  V   640

1921  TGCTGGCCCTCCGGCGAGGGAACGCTCAGTTGGCCGCACCTGCTCAGTGACCCGTCCATT  1980
 641  C  W  P  S  G  E  G  T  L  S  W  P  D  L  L  S  D  P  S  I   660

1981  GTGGGTAGCAATCTGCGGCAGCTGGCACGGGGCCAGGCGGGCCATGGGCTGGGCCCAGAG  2040
 661  V  G  S  N  L  R  Q  L  A  R  G  Q  A  G  H  G  L  G  P  E   680

2041  GAGGACGGCTTCTCCCTGGCCAGCCCCTCCTCGCCTGCCAAATCCTTCTCAGCATCAGAT  2100
 681  E  D  G  F  S  L  A  S  P  Y  S  P  A  K  S  F  S  A  S  D   700

2101  GAAGACCTGATCCAGGAGGTCCTTGCCGAGGGGGTCAGCAGCCCAGCCCCTACCCAAGAC  2160
 701  E  D  L  I  Q  Q  V  L  A  E  G  V  S  S  P  A  P  T  Q  D   720

2161  ACCCACATGGAAACGGACCTGCTCAGCAGCCTGTCCAGCACTCCTGGGGAGAAGACAGAG  2220
 721  T  H  M  E  T  D  L  L  S  S  L  S  S  T  P  G  E  K  T  E   740

2221  ACGCTGGCGCTGCAGAGGCTGGGGGAGCTGGGGCCACCCAGCCCAGGCCTGAACTGGGAA  2280
 741  T  L  A  L  Q  R  L  G  E  L  G  P  P  S  P  G  L  N  W  E   760

2281  CAGCCCCAGGCAGCGAGGCTGTCCAGGACAGGACTGGTGGAGGGTCTGCGGAAGCGCCTG  2340
 761  Q  P  Q  A  A  R  L  S  R  T  G  L  V  E  G  L  R  K  R  L   780

2341  CTGCCGGCCTGGTGTGCCTCCCTGGCCCACGGGCTCAGCCTGCTCCTGGTGGCTGTGGCT  2400
 781  L  P  A  W  C  A  S  L  A  H  G  L  S  L  L  L  V  A  V  A   800

2401  GTGGCTGTCTCAGGGTGGGTGGGTGCGAGCTTCCCCCCGGGCCTGAGTGTTGCGTGGCTC  2460
 801  V  A  V  S  G  W  V  G  A  S  F  P  P  G  V  S  V  A  W  L   820

2461  CTGTCCAGCAGCGCCAGCTTCCTGGCCTCATTCCTCGGCTGGGAGCCACTGAAGGTCTTG  2520
 821  L  S  S  S  A  S  F  L  A  S  F  L  G  W  E  P  L  K  V  L   840
```

Figure 7 cont'd

```
2521  CTGGAAGCCCTGTACTTCTCACTGGTGGCCAAGCGGCTGCACCCGGATGAAGATGACACC  2580
 841  L  E  A  L  Y  F  S  L  V  A  K  R  L  H  P  D  E  D  D  T    860

2581  CTGGTAGAGAGCCCGGCTGTGACGCCTGTGAGCGCACGTGTGCCCCGCGTACGGCCACCC  2640
 861  L  V  E  S  P  A  V  T  P  V  S  A  R  V  P  R  V  R  P  P    880

2641  CACGGCTTTGCACTCTTCCTGGCCAAGGAAGAAGCCCGCAAGGTCAAGAGGCTACATGGC  2700
 881  H  G  F  A  L  F  L  A  K  E  E  A  R  K  V  K  R  L  H  G    900

2701  ATGCTGCGGAGCCTCCTGGTGTACATGCTTTTTCTGCTGGTGACCCTGCTGGCCAGCTAT  2760
 901  M  L  R  S  L  L  V  Y  M  L  F  L  L  V  T  L  L  A  S  Y    920

2761  GGGGATGCCTCATGCCATGGGCACGCCTACCGTCTGCAAAGCGCCATCAAGCAGGAGCTG  2820
 921  G  D  A  S  C  H  G  H  A  Y  R  L  Q  S  A  I  K  Q  E  L    940

2821  CACAGCCGGGCCTTCCTGGCCATCACGCGGTCTGAGGAGCTCTGGCCATGGATGGCCCAC  2880
 941  H  S  R  A  F  L  A  I  T  R  S  E  E  L  W  P  W  M  A  H    960

2881  GTGCTGCTGCCCTACGTCCACGGGAACCAGTCCAGCCCAGAGCTGGGGCCCCCACGGCTG  2940
 961  V  L  L  P  Y  V  H  G  N  Q  S  S  P  E  L  G  P  P  R  L    980

2941  CGGCAGGTGCGGCTGCAGGAAGCACTCTACCCAGACCCTCCCGGCCCC`AGGGTCCACACG  3000
 981  R  Q  V  R  L  Q  E  A  L  Y  P  D  P  P  G  P  R  V  H  T   1000

3001  TGCTCGGCCGCAGGAGGCTTCAGCACCAGCGATTACGACGTTGGCTGGGAGAGTCCTCAC  3060
1001  C  S  A  A  G  G  F  S  T  S  D  Y  D  V  G  W  E  S  P  H   1020

3061  AATGGCTCGGGGACGTGGGCCTATTCAGCGCCGGATCTGCTGGGGGCATGGTCCTGGGGC  3120
1021  N  G  S  G  T  W  A  Y  S  A  P  D  L  L  G  A  W  S  W  G   1040

3121  TCCTGTGCCGTGTATGACAGCGGGGGCTACGTGCAGGAGCTGGGCCTGAGCCTGGAGGAG  3180
1041  S  C  A  V  Y  D  S  G  G  Y  V  Q  E  L  G  L  S  L  E  E   1060

3181  AGCCGCGACCGGCTGCGCTTCCTGCAGCTGCACAACTGGCTGGACAACAGGAGCCGCGCT  3240
1061  S  R  D  R  L  R  F  L  Q  L  H  N  W  L  D  N  R  S  R  A   1080

3241  GTGTTCCTGGAGCTCACGCGCTACAGCCCCGCCGTGGGGCTGCACGCCGCCGTCACGCTG  3300
1081  V  F  L  E  L  T  R  Y  S  P  A  V  G  L  H  A  A  V  T  L   1100

3301  CGCCTCGAGTTCCCCGGCGGCCGGCCGCGCCCTGGCCGCCCTCAGCGTCCGCCCCTTTGCG  3360
1101  R  L  E  F  P  A  A  G  R  A  L  A  A  L  S  V  R  P  F  A   1120

3361  CTGCGCCGCCTCAGCGCGGGCCTCTCGCTGCCTCTGCTCACCTCGGTGTGCCTGCTGCTG  3420
1121  L  R  R  L  S  A  G  L  S  L  P  L  L  T  S  V  C  L  L  L   1140

3421  TTCGCCGTGCACTTCGCCGTGGCCGAGGCCCGTACTTGGCACAGGGAAGGGCGCTGGCGC  3480
1141  F  A  V  H  F  A  V  A  E  A  R  T  W  H  R  E  G  R  W  R   1160

3481  GTGCTGCGGCTCGGAGCCTGGGCGCGGTGGCTGCTGGTGGCGCTGACGGCGGCCACGGCA  3540
1161  V  L  R  L  G  A  W  A  R  W  L  L  V  A  L  T  A  A  T  A   1180

3541  CTGGTACGCCTCGCCCAGCTGGGTGCGCTGACCGCCAGTGGACCCGTTTCGTGCGCGGC  3600
1181  L  V  R  L  A  Q  L  G  A  A  D  R  Q  W  T  R  F  V  R  G   1200

3601  CGCCCGCGCCGCTTCACTAGCTTCGACCAGGTGGCGCACGTGAGCTCCGCAGCCCGTGGC  3660
1201  R  P  R  R  F  T  S  F  D  Q  V  A  H  V  S  S  A  A  R  G   1220

3661  CTGGCGGCCTCGCTGCTCTTCCTGCTTTTGGTCAAGGCTGCCCAGCACGTACGCTTCGTG  3720
1221  L  A  A  S  L  L  F  L  L  L  V  K  A  A  Q  H  V  R  F  V   1240

3721  CGCCAGTGGTCCGTCTTTGGCAAGACATTATGCCGAGCTCTGCCAGAGCTCCTGGGGGTC  3780
1241  R  Q  W  S  V  F  G  K  T  L  C  R  A  L  P  E  L  L  G  V   1260
```

Figure 7 cont'd

```
3781  ACCTTGGGCCTGGTGGTGCTCGGGGTAGCCTACGCCCAGCTGGCCATCCTGCTCGTGTCT  3840
1261  T  L  G  L  V  V  L  G  V  A  Y  A  Q  L  A  I  L  L  V  S   1280

3841  TCCTGTGTGGACTCCCTCTGGAGCGTGGCCCAGGCCCTGTTGGTGCTGTGCCCTGGGACT  3900
1281  S  C  V  D  S  L  W  S  V  A  Q  A  L  L  V  L  C  P  G  T   1300

3901  GGGCTCTCTACCCTGTGTCCTGCCGAGTCCTGGCACCTGTCACCCCTGCTGTGTGTGGGG  3960
1301  G  L  S  T  L  C  P  A  E  S  W  H  L  S  P  L  L  C  V  G   1320

3961  CTCTGGGCACTGCGGCTGTGGGGCGCCCTACGGCTGGGGGCTGTTATTCTCCGCTGGCGC  4020
1321  L  W  A  L  R  L  W  G  A  L  R  L  G  A  V  I  L  R  W  R   1340

4021  TACCACGCCTTGCGTGGAGAGCTGTACCGGCCGGCCTGGGAGCCCCAGGACTACGAGATG  4080
1341  Y  H  A  L  R  G  E  L  Y  R  P  A  W  E  P  Q  D  Y  E  M   1360

4081  GTGGAGTTGTTCCTGCGCAGGCTGCGCCTCTGGATGGGCCTCAGCAAGGTCAAGGAGTTC  4140
1361  V  E  L  F  L  R  R  L  R  L  W  M  G  L  S  K  V  K  E  F   1380

4141  CGCCACAAAGTCCGCTTTGAAGGGATGGAGCCGCTGCCCTCTCGCTCCTCCAGGGGCTCC  4200
1381  R  H  K  V  R  F  E  G  M  E  P  L  P  S  R  S  S  R  G  S   1400

4201  AAGGTATCCCCGGATGTGCCCCCACCCAGCGCTGGCTCCGATGCCTCGCACCCCTCCACC  4260
1401  K  V  S  P  D  V  P  P  P  S  A  G  S  D  A  S  H  P  S  T   1420

4261  TCCTCCAGCCAGCTGGATGGGCTGAGCGTGAGCCTGGGCCGGCTGGGGACAAGGTGTGAG  4320
1421  S  S  S  Q  L  D  G  L  S  V  S  L  G  R  L  G  T  R  C  E   1440

4321  CCTGAGCCCTCCCGCCTCCAAGCCGTGTTCGAGGCCCTGCTCACCCAGTTTGACCGACTC  4380
1441  P  E  P  S  R  L  Q  A  V  F  E  A  L  L  T  Q  F  D  R  L   1460

4381  AACCAGGCCACAGAGGACGTCTACCAGCTGGAGCAGCAGCTGCACAGCCTGCAAGGCCGC  4440
1461  N  Q  A  T  E  D  V  Y  Q  L  E  Q  Q  L  H  S  L  Q  G  R   1480

4441  AGGAGCAGCCGGGCGCCCGCCGGATCTTCCCGTGGCCCATCCCCGGGCCTGCGGCCAGCA  4500
1481  R  S  S  R  A  P  A  G  S  S  R  G  P  S  P  G  L  R  P  A   1500

4501  CTGCCCAGCCGCCTTGCCCGGGCCAGTCGGGGTGTGGACCTGGCCACTGGCCCCAGCAGG  4560
1501  L  P  S  R  L  A  R  A  S  R  G  V  D  L  A  T  G  P  S  R   1520

4561  ACACCTTCGGGCCAAGAACAAGGTCCACCCCAGCAGCACTTAGTCCTCCTTCCTGGCGGG  4620
1521  T  P  S  G  Q  E  Q  G  P  P  Q  Q  H  L  V  L  L  P  G  G   1540

4621  GGTGGGCCGTGGAGTCGGAGTGGACACCGCTCAGTATTACTTTCTGCCGCTGTCAAGGCC  4689
                                                                     0
1541  G  G  P  W  S  R  S  G  H  R  S  V  L  L  S  A  A  V  K  A   1560

4681  GAGGGCCAGGCAGAATGGCTGCACGTAGGTTCCCCAGAGAGCAGGCAGGGGCATCTGTCT  4740
1561  E  G  Q  A  E  W  L  H  V  G  S  P  E  S  R  Q  G  H  L  S   1580

4741  GTCTGTGGGCTTCAGCACTTTAAAGAGGCTGTGTGGCCAACCAGGACCCAGGGTCCCCTC  4800
1581  V  C  G  L  Q  H  F  K  E  A  V  W  P  T  R  T  Q  G  P  L   1600

4801  CCCAGCTCCCTTGGGAAGGACACAGCAGTATTGGACGGTTTCTAGCCTCTGAGATGCTAA  4860
1601  P  S  S  L  G  K  D  T  A  V  L  D  G  F                     1620

4861  TTTATTTCCCCGAGTCCTCAGGTACAGCGGGCTGTGCCCGGCCCCACCCCCTGGGCAGAT  4920

4921  GTCCCCCCACTGCTAAGGCTGCTGGCTTCAGGGAGGGTTAGCCTGCACCGCCGCCACCCTG  4980

4981  CCCCTAAGTTATTACCTCTCCAGTTCCTACCGTACTCCCTGCACCGTCTCACTGTGTGTC  5040

5041  TCGTCTCAGTAATTTATATGGTGTTAAAATGTGTATATTTTTGTATGTCACTATTTTCAC  5100
```

Figure 7 Cont'd

```
5101  TAGGGCTGAGGGGCCTGCGCCCAGAGCTGGCCTCCCCCAACACCTGCTGCGCTTGGTAGG  5160
5161  TGTGGTGGCGTTATGGCAGCCCGGCTGCTGCTTGGATGCGAGCTTGGCCTTGGGCCGGTG  5220
5221  CTGGGGGCACAGCTGTCTGCCAGGCACTCTCATCACCCCAGAGGCCTTGTCATCCTCCCT  5280
5281  TGCCCCAGGCCAGGTAGCAAGAGAGCAGCGCCCAGGCCTGCTGGCATCAGGTCTGGGCAA  5340
5341  CTAGCAGGACTAGGCATGTCAGAGGACCCCAGGGTGGTTAGAGGAAAAGACTCCTCCTGG  5400
5401  GGGCTGGCTCCCAGGGTGGAGGAAGGTGACTGTGTGTGTGTGTGTGTGCGCGCGCGACGC  5460
5461  GCGACTGTGCTGTATGGCCCAGGCACGCTCAAGGCCCTCGGAGCTGGCTGTGCCTGCTTC  5520
5521  TGTGTACCACTTCTGTGGGCATGGCCGCTTCTAGAGCCTCGACACCCCCCCAACCCCCGC  5580
5581  ACCAAGCAGACAAAGTCAATAAAAGAGCTGTCTGACTGCAAAAAAAAAAAA 5631
```

1A1H0.6

```
   1  AAGCTTGGCA CCATCAAGGG CCAGTTCAAC TTTGTCCACG TGATCGTCAC CCCGCTGGAC
  61  TACGAGTGCA ACCTGGTGTC CCTGCAGTGC AGGAAAGACA TGGAGGGCCT TGTGGACACC
 121  AGCGTGGCCA AGATCGTGTC TGACCGCAAC CTGCCCTTCG TGGCCGCCA GATGGCCCTG
 181  CACGCAAATA TGGCCTCACA GGTGCATCAT AGCCGCTCCA ACCCCACCGA TATCTACCCC
 241  TCCAAGTGGA TTGCCCGGCT CCGCCACATC AAGCGGCTCC GCCAGCGGAT CTGCGAGGAA
 301  GCCGCCTACT CCAACCCCAG CCTACCTCTG GTGCACCCTC CGTCCCATAG CAAAGCCCCT
 361  GCACAGACTC CAGCCGAGCC CACACCTGGC TATGAGGTGG GCCAGCGGAA GCGCCTCATC
 421  TCCTCGGTGG AGGACTTCAC CGAGTTTGTG TGAGGCCGGG GCCCTCCCTC CTGCACTGGC
 481  CTTGGACGGT ATTGCCTGTC AGTGAAATAA ATAAAGTCCT GACCCCAGTG CACAGACATA
 541  GAGGCACAGA TTGC
```

Figure 8

WC10F

```
   1  GTCCGCGGTC GCACGTACGC TTCTGGTGTG TGTGAGACGT GCGGGGCTGG GAAGTGTTGG
  61  CAGACGGCGA GTACGTCCTC ACTCCTTTTG TTCTTTTGAC CTAAGCTGGC GAGTGGCACT
 121  GCTGAGTTCC GCTCAGTGCC CGCCCTGATG TGCGACCCCC GTGCATTCTT GCTGTTAGGT
 181  GGTGGCGGTG TG
```

CW10R

```
   1  AGGCAGGTCT CCCCCACGAG CAGGGGAGAG GCACCCAAGG T
```

Figure 9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1: (Compare Fig.1)

```
C GGC GCC GCC TGC CGC GTC AAC TGC TCG GGC CGC GGG CTG CGG ACG          46
  Gly Ala Ala Cys Arg Val Asn Cys Ser Gly Arg Gly Leu Arg Thr
    1               5                  10                  15

CTC GGT CCC GCG CTG CGC ATC CCC GCG GAC GCC ACA GCG CTA GAC GTC        94
Leu Gly Pro Ala Leu Arg Ile Pro Ala Asp Ala Thr Ala Leu Asp Val
                20                  25                  30

TCC CAC AAC CTG CTC CGG GCG CTG GAC GTT GGG CTC CTG GCG AAC CTC       142
Ser His Asn Leu Leu Arg Ala Leu Asp Val Gly Leu Leu Ala Asn Leu
            35                  40                  45

TCG GCG CTG GCA GAG CTG GAT ATA AGC AAC AAC AAG ATT TCT ACG TTA       190
Ser Ala Leu Ala Glu Leu Asp Ile Ser Asn Asn Lys Ile Ser Thr Leu
        50                  55                  60

GAA GAA GGA ATA TTT GCT AAT TTA TTT AAT TTA AGT GAA ATA AAC CTG       238
Glu Glu Gly Ile Phe Ala Asn Leu Phe Asn Leu Ser Glu Ile Asn Leu
    65                  70                  75

AGT GGG AAC CCG TTT GAG TGT GAC TGT GGC CTG GCG TGG CTG CCG CGA       286
Ser Gly Asn Pro Phe Glu Cys Asp Cys Gly Leu Ala Trp Leu Pro Arg
80                  85                  90                  95

TGG GCG GAG GAG CAG CAG GTG CGG GTG GTG CAG CCC GAG GCA GCC ACG       334
Trp Ala Glu Glu Gln Gln Val Arg Val Val Gln Pro Glu Ala Ala Thr
                100                 105                 110

TGT GCT GGG CCT GGC TCC CTG GCT GGC CAG CCT CTG CTT GGC ATC CCC       382
Cys Ala Gly Pro Gly Ser Leu Ala Gly Gln Pro Leu Leu Gly Ile Pro
            115                 120                 125

TTG CTG GAC AGT GGC TGT GGT GAG GAG TAT GTC GCC TGC CTC CCT GAC       430
Leu Leu Asp Ser Gly Cys Gly Glu Glu Tyr Val Ala Cys Leu Pro Asp
        130                 135                 140

AAC AGC TCA GGC ACC GTG GCA GCA GTG TCC TTT TCA GCT GCC CAC GAA       478
Asn Ser Ser Gly Thr Val Ala Ala Val Ser Phe Ser Ala Ala His Glu
    145                 150                 155

GGC CTG CTT CAG CCA GAG GCC TGC AGC GCC TTC TGC TTC TCC ACC GGC       526
Gly Leu Leu Gln Pro Glu Ala Cys Ser Ala Phe Cys Phe Ser Thr Gly
160                 165                 170                 175

CAG GGC CTC GCA GCC CTC TCG GAG CAG GGC TGG TGC CTG TGT GGG GCG       574
Gln Gly Leu Ala Ala Leu Ser Glu Gln Gly Trp Cys Leu Cys Gly Ala
                180                 185                 190

GCC CAG CCC TCC AGT GCC TCC TTT GCC TGC CTG TCC CTC TGC TCC GGC       622
Ala Gln Pro Ser Ser Ala Ser Phe Ala Cys Leu Ser Leu Cys Ser Gly
            195                 200                 205

CCC CCG CCA CCT CCT GCC CCC ACC TGT AGG GGC CCC ACC CTC CTC CAG       670
Pro Pro Pro Pro Pro Ala Pro Thr Cys Arg Gly Pro Thr Leu Leu Gln
        210                 215                 220

CAC GTC TTC CCT GCC TCC CCA GGG GCC ACC CTG GTG GGG CCC CAC GGA       718
His Val Phe Pro Ala Ser Pro Gly Ala Thr Leu Val Gly Pro His Gly
    225                 230                 235
```

```
CCT CTG GCC TCT GGC CAG CTA GCA GCC TTC CAC ATC GCT GCC CCG CTC          766
Pro Leu Ala Ser Gly Gln Leu Ala Ala Phe His Ile Ala Ala Pro Leu
240                 245                 250                 255

CCT GTC ACT GCC ACA CGC TGG GAC TTC GGA GAC GGC TCC GCC GAG GTG          814
Pro Val Thr Ala Thr Arg Trp Asp Phe Gly Asp Gly Ser Ala Glu Val
                260                 265                 270

GAT GCC GCT GGG CCG GCT GCC TCG CAT CGC TAT GTG CTG CCT GGG CGC          862
Asp Ala Ala Gly Pro Ala Ala Ser His Arg Tyr Val Leu Pro Gly Arg
            275                 280                 285

TAT CAC GTG ACG GCC GTG CTG GCC CTG GGG GCC GGC TCA GCC CTG CTG          910
Tyr His Val Thr Ala Val Leu Ala Leu Gly Ala Gly Ser Ala Leu Leu
        290                 295                 300

GGG ACA GAC GTG CAG GTG GAA GCG GCA CCT GCC GCC CTG GAG CTC GTG          958
Gly Thr Asp Val Gln Val Glu Ala Ala Pro Ala Ala Leu Glu Leu Val
    305                 310                 315

TGC CCG TCC TCG GTG CAG AGT GAC GAG AGC CTT GAC CTC AGC ATC CAG         1006
Cys Pro Ser Ser Val Gln Ser Asp Glu Ser Leu Asp Leu Ser Ile Gln
320                 325                 330                 335

AAC CGC GGT GGT TCA GGC CTG GAG GCC GCC TAC AGC ATC GTG GCC CTG         1054
Asn Arg Gly Gly Ser Gly Leu Glu Ala Ala Tyr Ser Ile Val Ala Leu
                340                 345                 350

GGC GAG GAG CCG GCC CGA GCG GTG CAC CCG CTC TGC CCC TCG GAC ACG         1102
Gly Glu Glu Pro Ala Arg Ala Val His Pro Leu Cys Pro Ser Asp Thr
            355                 360                 365

GAG ATC TTC CCT GGC AAC GGG CAC TGC TAC CGC CTG GTG GTG GAG AAG         1150
Glu Ile Phe Pro Gly Asn Gly His Cys Tyr Arg Leu Val Val Glu Lys
        370                 375                 380

GCG GCC TGG CTG CAG GCG CAG GAG CAG TGT CAG GCC TGG GCC GGG GCC         1198
Ala Ala Trp Leu Gln Ala Gln Glu Gln Cys Gln Ala Trp Ala Gly Ala
    385                 390                 395

GCC CTG GCA ATG GTG GAC AGT CCC GCC GTG CAG CGC TTC CTG GTC TCC         1246
Ala Leu Ala Met Val Asp Ser Pro Ala Val Gln Arg Phe Leu Val Ser
400                 405                 410                 415

CGG GTC ACC AGG AGC CTA GAC GTG TGG ATC GGC TTC TCG ACT GTG CAG         1294
Arg Val Thr Arg Ser Leu Asp Val Trp Ile Gly Phe Ser Thr Val Gln
                420                 425                 430

GGG GTG GAG GTG GGC CCA GCG CCG CAG GGC GAG GCC TTC AGC CTG GAG         1342
Gly Val Glu Val Gly Pro Ala Pro Gln Gly Glu Ala Phe Ser Leu Glu
            435                 440                 445

AGC TGC CAG AAC TGG CTG CCC GGG GAG CCA CAC CCA GCC ACA GCC GAG         1390
Ser Cys Gln Asn Trp Leu Pro Gly Glu Pro His Pro Ala Thr Ala Glu
        450                 455                 460

CAC TGC GTC CGG CTC GGG CCC ACC GGG TGG TGT AAC ACC GAC CTG TGC         1438
His Cys Val Arg Leu Gly Pro Thr Gly Trp Cys Asn Thr Asp Leu Cys
    465                 470                 475
```

```
TCA GCG CCG CAC AGC TAC GTC TGC GAG CTG CAG CCC GGA GGC CCA GTG      1486
Ser Ala Pro His Ser Tyr Val Cys Glu Leu Gln Pro Gly Gly Pro Val
480                 485                 490                 495

CAG GAT GCC GAG AAC CTC CTC GTG GGA GCG CCC AGT GGG GAC CTG CAG      1534
Gln Asp Ala Glu Asn Leu Leu Val Gly Ala Pro Ser Gly Asp Leu Gln
                500                 505                 510

GGA CCC CTG ACG CCT CTG GCA CAG CAG GAC GGC CTC TCA GCC CCG CAC      1582
Gly Pro Leu Thr Pro Leu Ala Gln Gln Asp Gly Leu Ser Ala Pro His
            515                 520                 525

GAG CCC GTG GAG GTC ATG GTA TTC CCG GGC CTG CGT CTG AGC CGT GAA      1630
Glu Pro Val Glu Val Met Val Phe Pro Gly Leu Arg Leu Ser Arg Glu
        530                 535                 540

GCC TTC CTC ACC ACG GCC GAA TTT GGG ACC CAG GAG CTC CGG CGG CCC      1678
Ala Phe Leu Thr Thr Ala Glu Phe Gly Thr Gln Glu Leu Arg Arg Pro
    545                 550                 555

GCC CAG CTG CGG CTG CAG GTG TAC CGG CTC CTC AGC ACA GCA GGG ACC      1726
Ala Gln Leu Arg Leu Gln Val Tyr Arg Leu Leu Ser Thr Ala Gly Thr
560                 565                 570                 575

CCG GAG AAC GGC AGC GAG CCT GAG AGC AGG TCC CCG GAC AAC AGG ACC      1774
Pro Glu Asn Gly Ser Glu Pro Glu Ser Arg Ser Pro Asp Asn Arg Thr
                580                 585                 590

CAG CTG GCC CCC GCG TGC ATG CCA GGG GGA CGC TGG TGC CCT GGA GCC      1822
Gln Leu Ala Pro Ala Cys Met Pro Gly Gly Arg Trp Cys Pro Gly Ala
            595                 600                 605

AAC ATC TGC TTG CCG CTG GAC GCC TCT TGC CAC CCC CAG GCC TGC GCC      1870
Asn Ile Cys Leu Pro Leu Asp Ala Ser Cys His Pro Gln Ala Cys Ala
        610                 615                 620

AAT GGC TGC ACG TCA GGG CCA GGG CTA CCC GGG GCC CCC TAT GCG CTA      1918
Asn Gly Cys Thr Ser Gly Pro Gly Leu Pro Gly Ala Pro Tyr Ala Leu
    625                 630                 635

TGG AGA GAG TTC CTC TTC TCC GTT GCC GCG GGG CCC CCC GCG CAG TAC      1966
Trp Arg Glu Phe Leu Phe Ser Val Ala Ala Gly Pro Pro Ala Gln Tyr
640                 645                 650                 655

TCG GTC ACC CTC CAC GGC CAG GAT GTC CTC ATG CTC CCT GGT GAC CTC      2014
Ser Val Thr Leu His Gly Gln Asp Val Leu Met Leu Pro Gly Asp Leu
                660                 665                 670

GTT GGC TTG CAG CAC GAC GCT GGC CCT GGC GCC CTC CTG CAC TGC TCG      2062
Val Gly Leu Gln His Asp Ala Gly Pro Gly Ala Leu Leu His Cys Ser
            675                 680                 685

CCG GCT CCC GGC CAC CCT GGT CCC CAG GCC CCG TAC CTC TCC GCC AAC      2110
Pro Ala Pro Gly His Pro Gly Pro Gln Ala Pro Tyr Leu Ser Ala Asn
        690                 695                 700

GCC TCG TCA TGG CTG CCC CAC TTG CCA GCC CAG CTG GAG GGC ACT TGG      2158
Ala Ser Ser Trp Leu Pro His Leu Pro Ala Gln Leu Glu Gly Thr Trp
    705                 710                 715
```

```
GCC TGC CCT GCC TGT GCC CTG CGG CTG CTT GCA GCC ACG GAA CAG CTC      2206
Ala Cys Pro Ala Cys Ala Leu Arg Leu Leu Ala Ala Thr Glu Gln Leu
720                 725                 730                 735

ACC GTG CTG CTG GGC TTG AGG CCC AAC CCT GGA CTG CGG ATG CCT GGG      2254
Thr Val Leu Leu Gly Leu Arg Pro Asn Pro Gly Leu Arg Met Pro Gly
                740                 745                 750

CGC TAT GAG GTC CGG GCA GAG GTG GGC AAT GGC GTG TCC AGG CAC AAC      2302
Arg Tyr Glu Val Arg Ala Glu Val Gly Asn Gly Val Ser Arg His Asn
            755                 760                 765

CTC TCC TGC AGC TTT GAC GTG GTC TCC CCA GTG GCT GGG CTG CGG GTC      2350
Leu Ser Cys Ser Phe Asp Val Val Ser Pro Val Ala Gly Leu Arg Val
        770                 775                 780

ATC TAC CCT GCC CCC CGC GAC GGC CGC CTC TAC GTG CCC ACC AAC GGC      2398
Ile Tyr Pro Ala Pro Arg Asp Gly Arg Leu Tyr Val Pro Thr Asn Gly
    785                 790                 795

TCA GCC TTG GTG CTC CAG GTG GAC TCT GGT GCC AAC GCC ACG GCC ACG      2446
Ser Ala Leu Val Leu Gln Val Asp Ser Gly Ala Asn Ala Thr Ala Thr
800                 805                 810                 815

GCT CGC TGG CCT GGG GGC AGT GTC AGC GCC CGC TTT GAG AAT GTC TGC      2494
Ala Arg Trp Pro Gly Gly Ser Val Ser Ala Arg Phe Glu Asn Val Cys
                820                 825                 830

CCT GCC CTG GTG GCC ACC TTC GTG CCC GGC TGC CCC TGG GAG ACC AAC      2542
Pro Ala Leu Val Ala Thr Phe Val Pro Gly Cys Pro Trp Glu Thr Asn
            835                 840                 845

GAT ACC CTG TTC TCA GTG GTA GCA CTG CCG TGG CTC AGT GAG GGG GAG      2590
Asp Thr Leu Phe Ser Val Val Ala Leu Pro Trp Leu Ser Glu Gly Glu
        850                 855                 860

CAC GTG GTG GAC GTG GTG GTG GAA AAC AGC GCC AGC CGG GCC AAC CTC      2638
His Val Val Asp Val Val Val Glu Asn Ser Ala Ser Arg Ala Asn Leu
    865                 870                 875

AGC CTG CGG GTG ACG GCG GAG GAG CCC ATC TGT GGC CTC CGC GCC ACG      2686
Ser Leu Arg Val Thr Ala Glu Glu Pro Ile Cys Gly Leu Arg Ala Thr
880                 885                 890                 895

CCC AGC CCC GAG GCC CGT GTA CTG CAG GGA GTC CTA GTG AGG TAC AGC      2734
Pro Ser Pro Glu Ala Arg Val Leu Gln Gly Val Leu Val Arg Tyr Ser
                900                 905                 910

CCC GTG GTG GAG GCC GGC TCG GAC ATG GTC TTC CGG TGG ACC ATC AAC      2782
Pro Val Val Glu Ala Gly Ser Asp Met Val Phe Arg Trp Thr Ile Asn
            915                 920                 925

GAC AAG CAG TCC CTG ACC TTC CAG AAC GTG GTC TTC AAT GTC ATT TAT      2830
Asp Lys Gln Ser Leu Thr Phe Gln Asn Val Val Phe Asn Val Ile Tyr
        930                 935                 940

CAG AGC GCG GCG GTC TTC AAG CTC TCA CTG ACG GCC TCC AAC CAC GTG      2878
Gln Ser Ala Ala Val Phe Lys Leu Ser Leu Thr Ala Ser Asn His Val
    945                 950                 955
```

```
AGC AAC GTC ACC GTG AAC TAC AAC GTA ACC GTG GAG CGG ATG AAC AGG      2926
Ser Asn Val Thr Val Asn Tyr Asn Val Thr Val Glu Arg Met Asn Arg
960                 965                 970                 975

ATG CAG GGT CTG CAG GTC TCC ACA GTG CCG GCC GTG CTG TCC CCC AAT      2974
Met Gln Gly Leu Gln Val Ser Thr Val Pro Ala Val Leu Ser Pro Asn
                980                 985                 990

GCC ACA CTG GTA CTG ACG GGT GGT GTG CTG GTG GAC TCA GCT GTG GAG      3022
Ala Thr Leu Val Leu Thr Gly Gly Val Leu Val Asp Ser Ala Val Glu
            995                 1000                1005

GTG GCC TTC CTG TGG AAC TTT GGG GAT GGG GAG CAG GCC CTC CAC CAG      3070
Val Ala Phe Leu Trp Asn Phe Gly Asp Gly Glu Gln Ala Leu His Gln
        1010                1015                1020

TTC CAG CCT CCG TAC AAC GAG TCC TTC CCG GTT CCA GAC CCC TCG GTG      3118
Phe Gln Pro Pro Tyr Asn Glu Ser Phe Pro Val Pro Asp Pro Ser Val
    1025                1030                1035

GCC CAG GTG CTG GTG GAG CAC AAT GTC ATG CAC ACC TAC GCT GCC CCA      3166
Ala Gln Val Leu Val Glu His Asn Val Met His Thr Tyr Ala Ala Pro
1040                1045                1050                1055

GGT GAG TAC CTC CTG ACC GTG CTG GCA TCT AAT GCC TTC GAG AAC CTG      3214
Gly Glu Tyr Leu Leu Thr Val Leu Ala Ser Asn Ala Phe Glu Asn Leu
                1060                1065                1070

ACG CAG CAG GTG CCT GTG AGC GTG CGC GCC TCC CTG CCC TCC GTG GCT      3262
Thr Gln Gln Val Pro Val Ser Val Arg Ala Ser Leu Pro Ser Val Ala
            1075                1080                1085

GTG GGT GTG AGT GAC GGC GTC CTG GTG GCC GGC CGG CCC GTC ACC TTC      3310
Val Gly Val Ser Asp Gly Val Leu Val Ala Gly Arg Pro Val Thr Phe
        1090                1095                1100

TAC CCG CAC CCG CTG CCC TCG CCT GGG GGT GTT CTT TAC ACG TGG GAC      3358
Tyr Pro His Pro Leu Pro Ser Pro Gly Gly Val Leu Tyr Thr Trp Asp
    1105                1110                1115

TTC GGG GAC GGC TCC CCT GTC CTG ACC CAG AGC CAG CCG GCT GCC AAC      3406
Phe Gly Asp Gly Ser Pro Val Leu Thr Gln Ser Gln Pro Ala Ala Asn
1120                1125                1130                1135

CAC ACC TAT GCC TCG AGG GGC ACC TAC CAC GTG CGC CTG GAG GTC AAC      3454
His Thr Tyr Ala Ser Arg Gly Thr Tyr His Val Arg Leu Glu Val Asn
                1140                1145                1150

AAC ACG GTG AGC GGT GCG GCG GCC CAG GCG GAT GTG CGC GTC TTT GAG      3502
Asn Thr Val Ser Gly Ala Ala Ala Gln Ala Asp Val Arg Val Phe Glu
            1155                1160                1165

GAG CTC CGC GGA CTC AGC GTG GAC ATG AGC CTG GCC GTG GAG CAG GGC      3550
Glu Leu Arg Gly Leu Ser Val Asp Met Ser Leu Ala Val Glu Gln Gly
        1170                1175                1180

GCC CCC GTG GTG GTC AGC GCC GCG GTG CAG ACG GGC GAC AAC ATC ACG      3598
Ala Pro Val Val Val Ser Ala Ala Val Gln Thr Gly Asp Asn Ile Thr
    1185                1190                1195
```

```
TGG ACC TTC GAC ATG GGG GAC GGC ACC GTG CTG TCG GGC CCG GAG GCA      3646
Trp Thr Phe Asp Met Gly Asp Gly Thr Val Leu Ser Gly Pro Glu Ala
1200                1205                1210                1215

ACA GTG GAG CAT GTG TAC CTG CGG GCA CAG AAC TGC ACA GTG ACC GTG      3694
Thr Val Glu His Val Tyr Leu Arg Ala Gln Asn Cys Thr Val Thr Val
                1220                1225                1230

GGT GCG GCC AGC CCC GCC GGC CAC CTG GCC CGG AGC CTG CAC GTG CTG      3742
Gly Ala Ala Ser Pro Ala Gly His Leu Ala Arg Ser Leu His Val Leu
            1235                1240                1245

GTC TTC GTC CTG GAG GTG CTG CGC GTT GAA CCC GCC GCC TGC ATC CCC      3790
Val Phe Val Leu Glu Val Leu Arg Val Glu Pro Ala Ala Cys Ile Pro
        1250                1255                1260

ACG CAG CCT GAC GCG CGG CTC ACG GCC TAC GTC ACC GGG AAC CCG GCC      3838
Thr Gln Pro Asp Ala Arg Leu Thr Ala Tyr Val Thr Gly Asn Pro Ala
    1265                1270                1275

CAC TAC CTC TTC GAC TGG ACC TTC GGG GAT GGC TCC TCC AAC ACG ACC      3886
His Tyr Leu Phe Asp Trp Thr Phe Gly Asp Gly Ser Ser Asn Thr Thr
1280                1285                1290                1295

GTG CGG GGG TGC CCG ACG GTG ACA CAC AAC TTC ACG CGG AGC GGC ACG      3934
Val Arg Gly Cys Pro Thr Val Thr His Asn Phe Thr Arg Ser Gly Thr
                1300                1305                1310

TTC CCC CTG GCG CTG GTG CTG TCC AGC CGC GTG AAC AGG GCG CAT TAC      3982
Phe Pro Leu Ala Leu Val Leu Ser Ser Arg Val Asn Arg Ala His Tyr
            1315                1320                1325

TTC ACC AGC ATC TGC GTG GAG CCA GAG GTG GGC AAC GTC ACC CTG CAG      4030
Phe Thr Ser Ile Cys Val Glu Pro Glu Val Gly Asn Val Thr Leu Gln
        1330                1335                1340

CCA GAG AGG CAG TTT GTG CAG CTC GGG GAC GAG GCC TGG CTG GTG GCA      4078
Pro Glu Arg Gln Phe Val Gln Leu Gly Asp Glu Ala Trp Leu Val Ala
    1345                1350                1355

TGT GCC TGG CCC CCG TTC CCC TAC CGC TAC ACC TGG GAC TTT GGC ACC      4126
Cys Ala Trp Pro Pro Phe Pro Tyr Arg Tyr Thr Trp Asp Phe Gly Thr
1360                1365                1370                1375

GAG GAA GCC GCC CCC ACC CGT GCC AGG GGC CCT GAG GTG ACG TTC ATC      4174
Glu Glu Ala Ala Pro Thr Arg Ala Arg Gly Pro Glu Val Thr Phe Ile
                1380                1385                1390

TAC CGA GAC CCA GGC TCC TAT CTT GTG ACA GTC ACC GCG TCC AAC AAC      4222
Tyr Arg Asp Pro Gly Ser Tyr Leu Val Thr Val Thr Ala Ser Asn Asn
            1395                1400                1405

ATC TCT GCT GCC AAT GAC TCA GCC CTG GTG GAG GTG CAG GAG CCC GTG      4270
Ile Ser Ala Ala Asn Asp Ser Ala Leu Val Glu Val Gln Glu Pro Val
        1410                1415                1420

CTG GTC ACC AGC ATC AAG GTC AAT GGC TCC CTT GGG CTG GAG CTG CAG      4318
Leu Val Thr Ser Ile Lys Val Asn Gly Ser Leu Gly Leu Glu Leu Gln
    1425                1430                1435
```

```
CAG CCG TAC CTG TTC TCT GCT GTG GGC CGT GGG CGC CCC GCC AGC TAC      4366
Gln Pro Tyr Leu Phe Ser Ala Val Gly Arg Gly Arg Pro Ala Ser Tyr
1440                 1445                 1450                 1455

CTG TGG GAT CTG GGG GAC GGT GGG TGG CTC GAG GGT CCG GAG GTC ACC      4414
Leu Trp Asp Leu Gly Asp Gly Gly Trp Leu Glu Gly Pro Glu Val Thr
                1460                 1465                 1470

CAC GCT TAC AAC AGC ACA GGT GAC TTC ACC GTT AGG GTG GCC GGC TGG      4462
His Ala Tyr Asn Ser Thr Gly Asp Phe Thr Val Arg Val Ala Gly Trp
            1475                 1480                 1485

AAT GAG GTG AGC CGC AGC GAG GCC TGG CTC AAT GTG ACG GTG AAG CGG      4510
Asn Glu Val Ser Arg Ser Glu Ala Trp Leu Asn Val Thr Val Lys Arg
        1490                 1495                 1500

CGC GTG CGG GGG CTC GTC GTC AAT GCA AGC CGC ACG GTG GTG CCC CTG      4558
Arg Val Arg Gly Leu Val Val Asn Ala Ser Arg Thr Val Val Pro Leu
    1505                 1510                 1515

AAT GGG AGC GTG AGC TTC AGC ACG TCG CTG GAG GCC GGC AGT GAT GTG      4606
Asn Gly Ser Val Ser Phe Ser Thr Ser Leu Glu Ala Gly Ser Asp Val
1520                 1525                 1530                 1535

CGC TAT TCC TGG GTG CTC TGT GAC CGC TGC ACG CCC ATC CCT GGG GGT      4654
Arg Tyr Ser Trp Val Leu Cys Asp Arg Cys Thr Pro Ile Pro Gly Gly
                1540                 1545                 1550

CCT ACC ATC TCT TAC ACC TTC CGC TCC GTG GGC ACC TTC AAT ATC ATC      4702
Pro Thr Ile Ser Tyr Thr Phe Arg Ser Val Gly Thr Phe Asn Ile Ile
            1555                 1560                 1565

GTC ACG GCT GAG AAC GAG GTG GGC TCC GCC CAG GAC AGC ATC TTC GTC      4750
Val Thr Ala Glu Asn Glu Val Gly Ser Ala Gln Asp Ser Ile Phe Val
        1570                 1575                 1580

TAT GTC CTG CAG CTC ATA GAG GGG CTG CAG GTG GTG GGC GGT GGC CGC      4798
Tyr Val Leu Gln Leu Ile Glu Gly Leu Gln Val Val Gly Gly Gly Arg
    1585                 1590                 1595

TAC TTC CCC ACC AAC CAC ACG GTA CAG CTG CAG GCC GTG GTT AGG GAT      4846
Tyr Phe Pro Thr Asn His Thr Val Gln Leu Gln Ala Val Val Arg Asp
1600                 1605                 1610                 1615

GGC ACC AAC GTC TCC TAC AGC TGG ACT GCC TGG AGG GAC AGG GGC CCG      4894
Gly Thr Asn Val Ser Tyr Ser Trp Thr Ala Trp Arg Asp Arg Gly Pro
                1620                 1625                 1630

GCC CTG GCC GGC AGC GGC AAA GGC TTC TCG CTC ACC GTG CTC GAG GCC      4942
Ala Leu Ala Gly Ser Gly Lys Gly Phe Ser Leu Thr Val Leu Glu Ala
            1635                 1640                 1645

GGC ACC TAC CAT GTG CAG CTG CGG GCC ACC AAC ATG CTG GGC AGC GCC      4990
Gly Thr Tyr His Val Gln Leu Arg Ala Thr Asn Met Leu Gly Ser Ala
        1650                 1655                 1660

TGG GCC GAC TGC ACC ATG GAC TTC GTG GAG CCT GTG GGG TGG CTG ATG      5038
Trp Ala Asp Cys Thr Met Asp Phe Val Glu Pro Val Gly Trp Leu Met
    1665                 1670                 1675
```

```
GTG ACC GCC TCC CCG AAC CCA GCT GCC GTC AAC ACA AGC GTC ACC CTC      5086
Val Thr Ala Ser Pro Asn Pro Ala Ala Val Asn Thr Ser Val Thr Leu
1680                 1685                 1690                 1695

AGT GCC GAG CTG GCT GGT GGC AGT GGT GTC GTA TAC ACT TGG TCC TTG      5134
Ser Ala Glu Leu Ala Gly Gly Ser Gly Val Val Tyr Thr Trp Ser Leu
                1700                 1705                 1710

GAG GAG GGG CTG AGC TGG GAG ACC TCC GAG CCA TTT ACC ACC CAT AGC      5182
Glu Glu Gly Leu Ser Trp Glu Thr Ser Glu Pro Phe Thr Thr His Ser
            1715                 1720                 1725

TTC CCC ACA CCC GGC CTG CAC TTG GTC ACC ATG ACG GCA GGG AAC CCG      5230
Phe Pro Thr Pro Gly Leu His Leu Val Thr Met Thr Ala Gly Asn Pro
        1730                 1735                 1740

CTG GGC TCA GCC AAC GCC ACC GTG GAA GTG GAT GTG CAG GTG CCT GTG      5278
Leu Gly Ser Ala Asn Ala Thr Val Glu Val Asp Val Gln Val Pro Val
    1745                 1750                 1755

AGT GGC CTC AGC ATC AGG GCC AGC GAG CCC GGA GGC AGC TTC GTG GCG      5326
Ser Gly Leu Ser Ile Arg Ala Ser Glu Pro Gly Gly Ser Phe Val Ala
1760                 1765                 1770                 1775

GCC GGG TCC TCT GTG CCC TTT TGG GGG CAG CTG GCC ACG GGC ACC AAT      5374
Ala Gly Ser Ser Val Pro Phe Trp Gly Gln Leu Ala Thr Gly Thr Asn
                1780                 1785                 1790

GTG AGC TGG TGC TGG GCT GTG CCC GGC GGC AGC AGC AAG CGT GGC CCT      5422
Val Ser Trp Cys Trp Ala Val Pro Gly Gly Ser Ser Lys Arg Gly Pro
            1795                 1800                 1805

CAT GTC ACC ATG GTC TTC CCG GAT GCT GGC ACC TTC TCC ATC CGG CTC      5470
His Val Thr Met Val Phe Pro Asp Ala Gly Thr Phe Ser Ile Arg Leu
        1810                 1815                 1820

AAT GCC TCC AAC GCA GTC AGC TGG GTC TCA GCC ACG TAC AAC CTC ACG      5518
Asn Ala Ser Asn Ala Val Ser Trp Val Ser Ala Thr Tyr Asn Leu Thr
    1825                 1830                 1835

GCG GAG GAG CCC ATC GTG GGC CTG GTG CTG TGG GCC AGC AGC AAG GTG      5566
Ala Glu Glu Pro Ile Val Gly Leu Val Leu Trp Ala Ser Ser Lys Val
1840                 1845                 1850                 1855

GTG GCG CCC GGG CAG CTG GTC CAT TTT CAG ATC CTG CTG GCT GCC GGC      5614
Val Ala Pro Gly Gln Leu Val His Phe Gln Ile Leu Leu Ala Ala Gly
                1860                 1865                 1870

TCA GCT GTC ACC TTC CGC CTG CAG GTC GGC GGG GCC AAC CCC GAG GTG      5662
Ser Ala Val Thr Phe Arg Leu Gln Val Gly Gly Ala Asn Pro Glu Val
            1875                 1880                 1885

CTC CCC GGG CCC CGT TTC TCC CAC AGC TTC CCC CGC GTC GGA GAC CAC      5710
Leu Pro Gly Pro Arg Phe Ser His Ser Phe Pro Arg Val Gly Asp His
        1890                 1895                 1900

GTG GTG AGC GTG CGG GGC AAA AAC CAC GTG AGC TGG GCC CAG GCG CAG      5758
Val Val Ser Val Arg Gly Lys Asn His Val Ser Trp Ala Gln Ala Gln
    1905                 1910                 1915
```

```
GTG CGC ATC GTG GTG CTG GAG GCC GTG AGT GGG CTG CAG ATG CCC AAC      5806
Val Arg Ile Val Val Leu Glu Ala Val Ser Gly Leu Gln Met Pro Asn
1920                1925                1930                1935

TGC TGC GAG CCT GGC ATC GCC ACG GGC ACT GAG AGG AAC TTC ACA GCC      5854
Cys Cys Glu Pro Gly Ile Ala Thr Gly Thr Glu Arg Asn Phe Thr Ala
                    1940                1945                1950

CGC GTG CAG CGC GGC TCT CGG GTC GCC TAC GCC TGG TAC TTC TCG CTG      5902
Arg Val Gln Arg Gly Ser Arg Val Ala Tyr Ala Trp Tyr Phe Ser Leu
            1955                1960                1965

CAG AAG GTC CAG GGC GAC TCG CTG GTC ATC CTG TCG GGC CGC GAC GTC      5950
Gln Lys Val Gln Gly Asp Ser Leu Val Ile Leu Ser Gly Arg Asp Val
        1970                1975                1980

ACC TAC ACG CCC GTG GCC GCG GGG CTG TTG GAG ATC CAG GTG CGC GCC      5998
Thr Tyr Thr Pro Val Ala Ala Gly Leu Leu Glu Ile Gln Val Arg Ala
    1985                1990                1995

TTC AAC GCC CTG GGC AGT GAG AAC CGC ACG CTG GTG CTG GAG GTT CAG      6046
Phe Asn Ala Leu Gly Ser Glu Asn Arg Thr Leu Val Leu Glu Val Gln
2000                2005                2010                2015

GAC GCC GTC CAG TAT GTG GCC CTG CAG AGC GGC CCC TGC TTC ACC AAC      6094
Asp Ala Val Gln Tyr Val Ala Leu Gln Ser Gly Pro Cys Phe Thr Asn
                    2020                2025                2030

CGC TCG GCG CAG TTT GAG GCC GCC ACC AGC CCC AGC CCC CGG CGT GTG      6142
Arg Ser Ala Gln Phe Glu Ala Ala Thr Ser Pro Ser Pro Arg Arg Val
            2035                2040                2045

GCC TAC CAC TGG GAC TTT GGG GAT GGG TCG CCA GGG CAG GAC ACA GAT      6190
Ala Tyr His Trp Asp Phe Gly Asp Gly Ser Pro Gly Gln Asp Thr Asp
        2050                2055                2060

GAG CCC AGG GCC GAG CAC TCC TAC CTG AGG CCT GGG GAC TAC CGC GTG      6238
Glu Pro Arg Ala Glu His Ser Tyr Leu Arg Pro Gly Asp Tyr Arg Val
    2065                2070                2075

CAG GTG AAC GCC TCC AAC CTG GTG AGC TTC TTC GTG GCG CAG GCC ACG      6286
Gln Val Asn Ala Ser Asn Leu Val Ser Phe Phe Val Ala Gln Ala Thr
2080                2085                2090                2095

GTG ACC GTC CAG GTG CTG GCC TGC CGG GAG CCG GAG GTG GAC GTG GTC      6334
Val Thr Val Gln Val Leu Ala Cys Arg Glu Pro Glu Val Asp Val Val
                    2100                2105                2110

CTG CCC CTG CAG GTG CTG ATG CGG CGA TCA CAG CGC AAC TAC TTG GAG      6382
Leu Pro Leu Gln Val Leu Met Arg Arg Ser Gln Arg Asn Tyr Leu Glu
            2115                2120                2125

GCC CAC GTT GAC CTG CGC GAC TGC GTC ACC TAC CAG ACT GAG TAC CGC      6430
Ala His Val Asp Leu Arg Asp Cys Val Thr Tyr Gln Thr Glu Tyr Arg
        2130                2135                2140

TGG GAG GTG TAT CGC ACC GCC AGC TGC CAG CGG CCG GGG CGC CCA GCG      6478
Trp Glu Val Tyr Arg Thr Ala Ser Cys Gln Arg Pro Gly Arg Pro Ala
    2145                2150                2155
```

```
CGT GTG GCC CTG CCC GGC GTG GAC GTG AGC CGG CCT CGG CTG GTG CTG      6526
Arg Val Ala Leu Pro Gly Val Asp Val Ser Arg Pro Arg Leu Val Leu
2160                2165                2170                2175

CCG CGG CTG GCG CTG CCT GTG GGG CAC TAC TGC TTT GTG TTT GTC GTG      6574
Pro Arg Leu Ala Leu Pro Val Gly His Tyr Cys Phe Val Phe Val Val
                    2180                2185                2190

TCA TTT GGG GAC ACG CCA CTG ACA CAG AGC ATC CAG GCC AAT GTG ACG      6622
Ser Phe Gly Asp Thr Pro Leu Thr Gln Ser Ile Gln Ala Asn Val Thr
            2195                2200                2205

GTG GCC CCC GAG CGC CTG GTG CCC ATC ATT GAG GGT GGC TCA TAC CGC      6670
Val Ala Pro Glu Arg Leu Val Pro Ile Ile Glu Gly Gly Ser Tyr Arg
        2210                2215                2220

GTG TGG TCA GAC ACA CGG GAC CTG GTG CTG GAT GGG AGC GAG TCC TAC      6718
Val Trp Ser Asp Thr Arg Asp Leu Val Leu Asp Gly Ser Glu Ser Tyr
    2225                2230                2235

GAC CCC AAC CTG GAG GAC GGC GAC CAG ACG CCG CTC AGT TTC CAC TGG      6766
Asp Pro Asn Leu Glu Asp Gly Asp Gln Thr Pro Leu Ser Phe His Trp
2240                2245                2250                2255

GCC TGT GTG GCT TCG ACA CAG AGG GAG GCT GGC GGG TGT GCG CTG AAC      6814
Ala Cys Val Ala Ser Thr Gln Arg Glu Ala Gly Gly Cys Ala Leu Asn
                    2260                2265                2270

TTT GGG CCC CGC GGG AGC AGC ACG GTC ACC ATT CCA CGG GAG CGG CTG      6862
Phe Gly Pro Arg Gly Ser Ser Thr Val Thr Ile Pro Arg Glu Arg Leu
            2275                2280                2285

GCG GCT GGC GTG GAG TAC ACC TTC AGC CTG ACC GTG TGG AAG GCC GGC      6910
Ala Ala Gly Val Glu Tyr Thr Phe Ser Leu Thr Val Trp Lys Ala Gly
        2290                2295                2300

CGC AAG GAG GAG GCC ACC AAC CAG ACG GTG CTG ATC CGG AGT GGC CGG      6958
Arg Lys Glu Glu Ala Thr Asn Gln Thr Val Leu Ile Arg Ser Gly Arg
    2305                2310                2315

GTG CCC ATT GTG TCC TTG GAG TGT GTG TCC TGC AAG GCA CAG GCC GTG      7006
Val Pro Ile Val Ser Leu Glu Cys Val Ser Cys Lys Ala Gln Ala Val
2320                2325                2330                2335

TAC GAA GTG AGC CGC AGC TCC TAC GTG TAC TTG GAG GGC CGC TGC CTC      7054
Tyr Glu Val Ser Arg Ser Ser Tyr Val Tyr Leu Glu Gly Arg Cys Leu
                    2340                2345                2350

AAT TGC AGC AGC GGC TCC AAG CGA GGG CGG TGG GCT GCA CGT ACG TTC      7102
Asn Cys Ser Ser Gly Ser Lys Arg Gly Arg Trp Ala Ala Arg Thr Phe
            2355                2360                2365

AGC AAC AAG ACG CTG GTG CTG GAT GAG ACC ACC ACA TCC ACG GGC AGT      7150
Ser Asn Lys Thr Leu Val Leu Asp Glu Thr Thr Tnr Ser Thr Gly Ser
        2370                2375                2380

GCA GGC ATG CGA CTG GTG CTG CGG CGG GGC GTG CTG CGG GAC GGC GAG      7198
Ala Gly Met Arg Leu Val Leu Arg Arg Gly Val Leu Arg Asp Gly Glu
    2385                2390                2395
```

```
GGA TAC ACC TTC ACG CTC ACG GTG CTG GGC CGC TCT GGC GAG GAG GAG      7246
Gly Tyr Thr Phe Thr Leu Thr Val Leu Gly Arg Ser Gly Glu Glu Glu
2400                2405                2410                2415

GGC TGC GCC TCC ATC CGC CTG TCC CCC AAC CGC CCG CCG CTG GGG GGC      7294
Gly Cys Ala Ser Ile Arg Leu Ser Pro Asn Arg Pro Pro Leu Gly Gly
                    2420                2425                2430

TCT TGC CGC CTC TTC CCA CTG GGC GCT GTG CAC GCC CTC ACC ACC AAG      7342
Ser Cys Arg Leu Phe Pro Leu Gly Ala Val His Ala Leu Thr Thr Lys
            2435                2440                2445

GTG CAC TTC GAA TGC ACG GGC TGG CAT GAC GCG GAG GAT GCT GGC GCC      7390
Val His Phe Glu Cys Thr Gly Trp His Asp Ala Glu Asp Ala Gly Ala
        2450                2455                2460

CCG CTG GTG TAC GCC CTG CTG CTG CGG CGC TGT CGC CAG GGC CAC TGC      7438
Pro Leu Val Tyr Ala Leu Leu Leu Arg Arg Cys Arg Gln Gly His Cys
    2465                2470                2475

GAG GAG TTC TGT GTC TAC AAG GGC AGC CTC TCC AGC TAC GGA GCC GTG      7486
Glu Glu Phe Cys Val Tyr Lys Gly Ser Leu Ser Ser Tyr Gly Ala Val
2480                2485                2490                2495

CTG CCC CCG GGT TTC AGG CCA CAC TTC GAG GTG GGC CTG GCC GTG GTG      7534
Leu Pro Pro Gly Phe Arg Pro His Phe Glu Val Gly Leu Ala Val Val
                    2500                2505                2510

GTG CAG GAC CAG CTG GGA GCC GCT GTG GTC GCC CTC AAC AGG TCT TTG      7582
Val Gln Asp Gln Leu Gly Ala Ala Val Val Ala Leu Asn Arg Ser Leu
            2515                2520                2525

GCC ATC ACC CTC CCA GAG CCC AAC GGC AGC GCA ACG GGG CTC ACA GTC      7630
Ala Ile Thr Leu Pro Glu Pro Asn Gly Ser Ala Thr Gly Leu Thr Val
        2530                2535                2540

TGG CTG CAC GGG CTC ACC GCT AGT GTG CTC CCA GGG CTG CTG CGG CAG      7678
Trp Leu His Gly Leu Thr Ala Ser Val Leu Pro Gly Leu Leu Arg Gln
    2545                2550                2555

GCC GAT CCC CAG CAC GTC ATC GAG TAC TCG TTG GCC CTG GTC ACC GTG      7726
Ala Asp Pro Gln His Val Ile Glu Tyr Ser Leu Ala Leu Val Thr Val
2560                2565                2570                2575

CTG AAC GAG TAC GAG CGG GCC CTG GAC GTG GCG GCA GAG CCC AAG CAC      7774
Leu Asn Glu Tyr Glu Arg Ala Leu Asp Val Ala Ala Glu Pro Lys His
                    2580                2585                2590

GAG CGG CAG CAC CGA GCC CAG ATA CGC AAG AAC ATC ACG GAG ACT CTG      7822
Glu Arg Gln His Arg Ala Gln Ile Arg Lys Asn Ile Thr Glu Thr Leu
            2595                2600                2605

GTG TCC CTG AGG GTC CAC ACT GTG GAT GAC ATC CAG CAG ATC GCT GCT      7870
Val Ser Leu Arg Val His Thr Val Asp Asp Ile Gln Gln Ile Ala Ala
        2610                2615                2620

GCG CTG GCC CAG TGC ATG GGG CCC AGC AGG GAG CTC GTA TGC CGC TCG      7918
Ala Leu Ala Gln Cys Met Gly Pro Ser Arg Glu Leu Val Cys Arg Ser
    2625                2630                2635
```

```
TGC CTG AAG CAG ACG CTG CAC AAG CTG GAG GCC ATG ATG CTC ATC CTG      7966
Cys Leu Lys Gln Thr Leu His Lys Leu Glu Ala Met Met Leu Ile Leu
2640                 2645                 2650                 2655

CAG GCA GAG ACC ACC GCG GGC ACC GTG ACG CCC ACC GCC ATC GGA GAC      8014
Gln Ala Glu Thr Thr Ala Gly Thr Val Thr Pro Thr Ala Ile Gly Asp
                2660                 2665                 2670

AGC ATC CTC AAC ATC ACA GGA GAC CTC ATC CAC CTG GCC AGC TCG GAC      8062
Ser Ile Leu Asn Ile Thr Gly Asp Leu Ile His Leu Ala Ser Ser Asp
            2675                 2680                 2685

GTG CGG GCA CCA CAG CCC TCA GAG CTG GGA GCC GAG TCA CCA TCT CGG      8110
Val Arg Ala Pro Gln Pro Ser Glu Leu Gly Ala Glu Ser Pro Ser Arg
        2690                 2695                 2700

ATG GTG GCG TCC CAG GCC TAC AAC CTG ACC TCT GCC CTC ATG CGC ATC      8158
Met Val Ala Ser Gln Ala Tyr Asn Leu Thr Ser Ala Leu Met Arg Ile
    2705                 2710                 2715

CTC ATG CGC TCC CGC GTG CTC AAC GAG GAG CCC CTG ACG CTG GCG GGC      8206
Leu Met Arg Ser Arg Val Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly
2720                 2725                 2730                 2735

GAG GAG ATC GTG GCC CAG GGC AAG CGC TCG GAC CCG CGG AGC CTG CTG      8254
Glu Glu Ile Val Ala Gln Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu
                2740                 2745                 2750

TGC TAT GGC GGC GCC CCA GGG CCT GGC TGC CAC TTC TCC ATC CCC GAG      8302
Cys Tyr Gly Gly Ala Pro Gly Pro Gly Cys His Phe Ser Ile Pro Glu
            2755                 2760                 2765

GCT TTC AGC GGG GCC CTG GCC AAC CTC AGT GAC GTG GTG CAG CTC ATC      8350
Ala Phe Ser Gly Ala Leu Ala Asn Leu Ser Asp Val Val Gln Leu Ile
        2770                 2775                 2780

TTT CTG GTG GAC TCC AAT CCC TTT CCC TTT GGC TAT ATC AGC AAC TAC      8398
Phe Leu Val Asp Ser Asn Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr
    2785                 2790                 2795

ACC GTC TCC ACC AAG GTG GCC TCG ATG GCA TTC CAG ACA CAG GCC GGC      8446
Thr Val Ser Thr Lys Val Ala Ser Met Ala Phe Gln Thr Gln Ala Gly
2800                 2805                 2810                 2815

GCC CAG ATC CCC ATC GAG CGG CTG GCC TCA GAG CGC GCC ATC ACC GTG      8494
Ala Gln Ile Pro Ile Glu Arg Leu Ala Ser Glu Arg Ala Ile Thr Val
                2820                 2825                 2830

AAG GTG CCC AAC AAC TCG GAC TGG GCT GCC CGG GGC CAC CGC AGC TCC      8542
Lys Val Pro Asn Asn Ser Asp Trp Ala Ala Arg Gly His Arg Ser Ser
            2835                 2840                 2845

GCC AAC TCC GCC AAC TCC GTT GTG GTC CAG CCC CAG GCC TCC GTC GGT      8590
Ala Asn Ser Ala Asn Ser Val Val Val Gln Pro Gln Ala Ser Val Gly
        2850                 2855                 2860

GCT GTG GTC ACC CTG GAC AGC AGC AAC CCT GCG GCC GGG CTG CAT CTG      8638
Ala Val Val Thr Leu Asp Ser Ser Asn Pro Ala Ala Gly Leu His Leu
    2865                 2870                 2875
```

```
CAG CTC AAC TAT ACG CTG CTG GAC GGC CAC TAC CTG TCT GAG GAA CCT      8686
Gln Leu Asn Tyr Thr Leu Leu Asp Gly His Tyr Leu Ser Glu Glu Pro
2880                2885                2890                2895

GAG CCC TAC CTG GCA GTC TAC CTA CAC TCG GAG CCC CGG CCC AAT GAG      8734
Glu Pro Tyr Leu Ala Val Tyr Leu His Ser Glu Pro Arg Pro Asn Glu
                    2900                2905                2910

CAC AAC TGC TCG GCT AGC AGG AGG ATC CGC CCA GAG TCA CTC CAG GGT      8782
His Asn Cys Ser Ala Ser Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly
            2915                2920                2925

GCT GAC CAC CGG CCC TAC ACC TTC TTC ATT TCC CCG GGG AGC AGA GAC      8830
Ala Asp His Arg Pro Tyr Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp
        2930                2935                2940

CCA GCG GGG AGT TAC CAT CTG AAC CTC TCC AGC CAC TTC CGC TGG TCG      8878
Pro Ala Gly Ser Tyr His Leu Asn Leu Ser Ser His Phe Arg Trp Ser
    2945                2950                2955

GCG CTG CAG GTG TCC GTG GGC CTG TAC ACG TCC CTG TGC CAG TAC TTC      8926
Ala Leu Gln Val Ser Val Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe
2960                2965                2970                2975

AGC GAG GAG GAC ATG GTG TGG CGG ACA GAG GGG CTG CTG CCC CTG GAG      8974
Ser Glu Glu Asp Met Val Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu
                    2980                2985                2990

GAG ACC TCG CCC CGC CAG GCC GTC TGC CTC ACC CGC CAC CTC ACC GCC      9022
Glu Thr Ser Pro Arg Gln Ala Val Cys Leu Thr Arg His Leu Thr Ala
            2995                3000                3005

TTC GGC GCC AGC CTC TTC GTG CCC CCA AGC CAT GTC CGC TTT GTG TTT      9070
Phe Gly Ala Ser Leu Phe Val Pro Pro Ser His Val Arg Phe Val Phe
        3010                3015                3020

CCT GAG CCG ACA GCG GAT GTA AAC TAC ATC GTC ATG CTG ACA TGT GCT      9118
Pro Glu Pro Thr Ala Asp Val Asn Tyr Ile Val Met Leu Thr Cys Ala
    3025                3030                3035

GTG TGC CTG GTG ACC TAC ATG GTC ATG GCC GCC ATC CTG CAC AAG CTG      9166
Val Cys Leu Val Thr Tyr Met Val Met Ala Ala Ile Leu His Lys Leu
3040                3045                3050                3055

GAC CAG TTG GAT GCC AGC CGG GGC CGC GCC ATC CCT TTC TGT GGG CAG      9214
Asp Gln Leu Asp Ala Ser Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln
                    3060                3065                3070

CGG GGC CGC TTC AAG TAC GAG ATC CTC GTC AAG ACA GGC TGG GGC CGG      9262
Arg Gly Arg Phe Lys Tyr Glu Ile Leu Val Lys Thr Gly Trp Gly Arg
            3075                3080                3085

GGC TCA GGT ACC ACG GCC CAC GTG GGC ATC ATG CTG TAT GGG GTG GAC      9310
Gly Ser Gly Thr Thr Ala His Val Gly Ile Met Leu Tyr Gly Val Asp
        3090                3095                3100

AGC CGG AGC GGC CAC CGG CAC CTG GAC GGC GAC AGA GCC TTC CAC CGC      9358
Ser Arg Ser Gly His Arg His Leu Asp Gly Asp Arg Ala Phe His Arg
    3105                3110                3115
```

```
AAC AGC CTG GAC ATC TTC CGG ATC GCC ACC CCG CAC AGC CTG GGT AGC          9406
Asn Ser Leu Asp Ile Phe Arg Ile Ala Thr Pro His Ser Leu Gly Ser
3120                 3125                 3130                 3135

GTG TGG AAG ATC CGA GTG TGG CAC GAC AAC AAA GGG CTC AGC CCT GCC          9454
Val Trp Lys Ile Arg Val Trp His Asp Asn Lys Gly Leu Ser Pro Ala
                3140                 3145                 3150

TGG TTC CTG CAG CAC GTC ATC GTC AGG GAC CTG CAG ACG GCA CGC AGC          9502
Trp Phe Leu Gln His Val Ile Val Arg Asp Leu Gln Thr Ala Arg Ser
            3155                 3160                 3165

GCC TTC TTC CTG GTC AAT GAC TGG CTT TCG GTG GAG ACG GAG GCC AAC          9550
Ala Phe Phe Leu Val Asn Asp Trp Leu Ser Val Glu Thr Glu Ala Asn
        3170                 3175                 3180

GGG GGC CTG GTG GAG AAG GAG GTG CTG GCC GCG AGC GAC GCA GCC CTT          9598
Gly Gly Leu Val Glu Lys Glu Val Leu Ala Ala Ser Asp Ala Ala Leu
    3185                 3190                 3195

TTG CGC TTC CGG CGC CTG CTG GTG GCT GAG CTG CAG CGT GGC TTC TTT          9646
Leu Arg Phe Arg Arg Leu Leu Val Ala Glu Leu Gln Arg Gly Phe Phe
3200                 3205                 3210                 3215

GAC AAG CAC ATC TGG CTC TCC ATA TGG GAC CGG CCG CCT CGT AGC CGT          9694
Asp Lys His Ile Trp Leu Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg
                3220                 3225                 3230

TTC ACT CGC ATC CAG AGG GCC ACC TGC TGC GTT CTC CTC ATC TGC CTC          9742
Phe Thr Arg Ile Gln Arg Ala Thr Cys Cys Val Leu Leu Ile Cys Leu
            3235                 3240                 3245

TTC CTG GGC GCC AAC GCC GTG TGG TAC GGG GCT GTT GGC GAC TCT GCC          9790
Phe Leu Gly Ala Asn Ala Val Trp Tyr Gly Ala Val Gly Asp Ser Ala
        3250                 3255                 3260

TAC AGC ACG GGG CAT GTG TCC AGG CTG AGC CCG CTG AGC GTC GAC ACA          9838
Tyr Ser Thr Gly His Val Ser Arg Leu Ser Pro Leu Ser Val Asp Thr
    3265                 3270                 3275

GTC GCT GTT GGC CTG GTG TCC AGC GTG GTT GTC TAT CCC GTC TAC CTG          9886
Val Ala Val Gly Leu Val Ser Ser Val Val Val Tyr Pro Val Tyr Leu
3280                 3285                 3290                 3295

GCC ATC CTT TTT CTC TTC CGG ATG TCC CGG AGC AAG GTG GCT GGG AGC          9934
Ala Ile Leu Phe Leu Phe Arg Met Ser Arg Ser Lys Val Ala Gly Ser
                3300                 3305                 3310

CCG AGC CCC ACA CCT GCC GGG CAG CAG GTG CTG GAC ATC GAC AGC TGC          9982
Pro Ser Pro Thr Pro Ala Gly Gln Gln Val Leu Asp Ile Asp Ser Cys
            3315                 3320                 3325

CTG GAC TCG TCC GTG CTG GAC AGC TCC TTC CTC ACG TTC TCA GGC CTC         10030
Leu Asp Ser Ser Val Leu Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu
        3330                 3335                 3340

CAC GCT GAG GCC TTT GTT GGA CAG ATG AAG AGT GAC TTG TTT CTG GAT         10078
His Ala Glu Ala Phe Val Gly Gln Met Lys Ser Asp Leu Phe Leu Asp
    3345                 3350                 3355
```

```
GAT TCT AAG AGT CTG GTG TGC TGG CCC TCC GGC GAG GGA ACG CTC AGT      10126
Asp Ser Lys Ser Leu Val Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser
3360                 3365                 3370                 3375

TGG CCG GAC CTG CTC AGT GAC CCG TCC ATT GTG GGT AGC AAT CTG CGG      10174
Trp Pro Asp Leu Leu Ser Asp Pro Ser Ile Val Gly Ser Asn Leu Arg
                3380                 3385                 3390

CAG CTG GCA CGG GGC CAG GCG GGC CAT GGG CTG GGC CCA GAG GAG GAC      10222
Gln Leu Ala Arg Gly Gln Ala Gly His Gly Leu Gly Pro Glu Glu Asp
            3395                 3400                 3405

GGC TTC TCC CTG GCC AGC CCC TAC TCG CCT GCC AAA TCC TTC TCA GCA      10270
Gly Phe Ser Leu Ala Ser Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala
        3410                 3415                 3420

TCA GAT GAA GAC CTG ATC CAG CAG GTC CTT GCC GAG GGG GTC AGC AGC      10318
Ser Asp Glu Asp Leu Ile Gln Gln Val Leu Ala Glu Gly Val Ser Ser
    3425                 3430                 3435

CCA GCC CCT ACC CAA GAC ACC CAC ATG GAA ACG GAC CTG CTC AGC AGC      10366
Pro Ala Pro Thr Gln Asp Thr His Met Glu Thr Asp Leu Leu Ser Ser
3440                 3445                 3450                 3455

CTG TCC AGC ACT CCT GGG GAG AAG ACA GAG ACG CTG GCG CTG CAG AGG      10414
Leu Ser Ser Thr Pro Gly Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg
                3460                 3465                 3470

CTG GGG GAG CTG GGG CCA CCC AGC CCA GGC CTG AAC TGG GAA CAG CCC      10462
Leu Gly Glu Leu Gly Pro Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro
            3475                 3480                 3485

CAG GCA GCG AGG CTG TCC AGG ACA GGA CTG GTG GAG GGT CTG CGG AAG      10510
Gln Ala Ala Arg Leu Ser Arg Thr Gly Leu Val Glu Gly Leu Arg Lys
        3490                 3495                 3500

CGC CTG CTG CCG GCC TGG TGT GCC TCC CTG GCC CAC GGG CTC AGC CTG      10558
Arg Leu Leu Pro Ala Trp Cys Ala Ser Leu Ala His Gly Leu Ser Leu
    3505                 3510                 3515

CTC CTG GTG GCT GTG GCT GTG GCT GTC TCA GGG TGG GTG GGT GCG AGC      10606
Leu Leu Val Ala Val Ala Val Ala Val Ser Gly Trp Val Gly Ala Ser
3520                 3525                 3530                 3535

TTC CCC CCG GGC GTG AGT GTT GCG TGG CTC CTG TCC AGC AGC GCC AGC      10654
Phe Pro Pro Gly Val Ser Val Ala Trp Leu Leu Ser Ser Ser Ala Ser
                3540                 3545                 3550

TTC CTG GCC TCA TTC CTC GGC TGG GAG CCA CTG AAG GTC TTG CTG GAA      10702
Phe Leu Ala Ser Phe Leu Gly Trp Glu Pro Leu Lys Val Leu Leu Glu
            3555                 3560                 3565

GCC CTG TAC TTC TCA CTG GTG GCC AAG CGG CTG CAC CCG GAT GAA GAT      10750
Ala Leu Tyr Phe Ser Leu Val Ala Lys Arg Leu His Pro Asp Glu Asp
        3570                 3575                 3580

GAC ACC CTG GTA GAG AGC CCG GCT GTG ACG CCT GTG AGC GCA CGT GTG      10798
Asp Thr Leu Val Glu Ser Pro Ala Val Thr Pro Val Ser Ala Arg Val
    3585                 3590                 3595
```

```
CCC CGC GTA CGG CCA CCC CAC GGC TTT GCA CTC TTC CTG GCC AAG GAA      10846
Pro Arg Val Arg Pro Pro His Gly Phe Ala Leu Phe Leu Ala Lys Glu
3600                 3605                 3610                 3615

GAA GCC CGC AAG GTC AAG AGG CTA CAT GGC ATG CTG CGG AGC CTC CTG      10894
Glu Ala Arg Lys Val Lys Arg Leu His Gly Met Leu Arg Ser Leu Leu
                3620                 3625                 3630

GTG TAC ATG CTT TTT CTG CTG GTG ACC CTG CTG GCC AGC TAT GGG GAT      10942
Val Tyr Met Leu Phe Leu Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp
            3635                 3640                 3645

GCC TCA TGC CAT GGG CAC GCC TAC CGT CTG CAA AGC GCC ATC AAG CAG      10990
Ala Ser Cys His Gly His Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln
        3650                 3655                 3660

GAG CTG CAC AGC CGG GCC TTC CTG GCC ATC ACG CGG TCT GAG GAG CTC      11038
Glu Leu His Ser Arg Ala Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu
    3665                 3670                 3675

TGG CCA TGG ATG GCC CAC GTG CTG CTG CCC TAC GTC CAC GGG AAC CAG      11086
Trp Pro Trp Met Ala His Val Leu Leu Pro Tyr Val His Gly Asn Gln
3680                 3685                 3690                 3695

TCC AGC CCA GAG CTG GGG CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG      11134
Ser Ser Pro Glu Leu Gly Pro Pro Arg Leu Arg Gln Val Arg Leu Gln
                3700                 3705                 3710

GAA GCA CTC TAC CCA GAC CCT CCC GGC CCC AGG GTC CAC ACG TGC TCG      11182
Glu Ala Leu Tyr Pro Asp Pro Pro Gly Pro Arg Val His Thr Cys Ser
            3715                 3720                 3725

GCC GCA GGA GGC TTC AGC ACC AGC GAT TAC GAC GTT GGC TGG GAG AGT      11230
Ala Ala Gly Gly Phe Ser Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser
        3730                 3735                 3740

CCT CAC AAT GGC TCG GGG ACG TGG GCC TAT TCA GCG CCG GAT CTG CTG      11278
Pro His Asn Gly Ser Gly Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu
    3745                 3750                 3755

GGG GCA TGG TCC TGG GGC TCC TGT GCC GTG TAT GAC AGC GGG GGC TAC      11326
Gly Ala Trp Ser Trp Gly Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr
3760                 3765                 3770                 3775

GTG CAG GAG CTG GGC CTG AGC CTG GAG GAG AGC CGC GAC CGG CTG CGC      11374
Val Gln Glu Leu Gly Leu Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg
                3780                 3785                 3790

TTC CTG CAG CTG CAC AAC TGG CTG GAC AAC AGG AGC CGC GCT GTG TTC      11422
Phe Leu Gln Leu His Asn Trp Leu Asp Asn Arg Ser Arg Ala Val Phe
            3795                 3800                 3805

CTG GAG CTC ACG CGC TAC AGC CCG GCC GTG GGG CTG CAC GCC GCC GTC      11470
Leu Glu Leu Thr Arg Tyr Ser Pro Ala Val Gly Leu His Ala Ala Val
        3810                 3815                 3820

ACG CTG CGC CTC GAG TTC CCG GCG GCC GGC CGC GCC CTG GCC GCC CTC      11518
Thr Leu Arg Leu Glu Phe Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu
    3825                 3830                 3835
```

```
AGC GTC CGC CCC TTT GCG CTG CGC CGC CTC AGC GCG GGC CTC TCG CTG      11566
Ser Val Arg Pro Phe Ala Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu
3840                 3845                 3850                 3855

CCT CTG CTC ACC TCG GTG TGC CTG CTG CTG TTC GCC GTG CAC TTC GCC      11614
Pro Leu Leu Thr Ser Val Cys Leu Leu Leu Phe Ala Val His Phe Ala
                3860                 3865                 3870

GTG GCC GAG GCC CGT ACT TGG CAC AGG GAA GGG CGC TGG CGC GTG CTG      11662
Val Ala Glu Ala Arg Thr Trp His Arg Glu Gly Arg Trp Arg Val Leu
            3875                 3880                 3885

CGG CTC GGA GCC TGG GCG CGG TGG CTG CTG GTG GCG CTG ACG GCG GCC      11710
Arg Leu Gly Ala Trp Ala Arg Trp Leu Leu Val Ala Leu Thr Ala Ala
        3890                 3895                 3900

ACG GCA CTG GTA CGC CTC GCC CAG CTG GGT GCC GCT GAC CGC CAG TGG      11758
Thr Ala Leu Val Arg Leu Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp
    3905                 3910                 3915

ACC CGT TTC GTG CGC GGC CGC CCG CGC CGC TTC ACT AGC TTC GAC CAG      11806
Thr Arg Phe Val Arg Gly Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln
3920                 3925                 3930                 3935

GTG GCG CAC GTG AGC TCC GCA GCC CGT GGC CTG GCG GCC TCG CTG CTC      11854
Val Ala His Val Ser Ser Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu
                3940                 3945                 3950

TTC CTG CTT TTG GTC AAG GCT GCC CAG CAC GTA CGC TTC GTG CGC CAG      11902
Phe Leu Leu Leu Val Lys Ala Ala Gln His Val Arg Phe Val Arg Gln
            3955                 3960                 3965

TGG TCC GTC TTT GGC AAG ACA TTA TGC CGA GCT CTG CCA GAG CTC CTG      11950
Trp Ser Val Phe Gly Lys Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu
        3970                 3975                 3980

GGG GTC ACC TTG GGC CTG GTG GTG CTC GGG GTA GCC TAC GCC CAG CTG      11998
Gly Val Thr Leu Gly Leu Val Val Leu Gly Val Ala Tyr Ala Gln Leu
    3985                 3990                 3995

GCC ATC CTG CTC GTG TCT TCC TGT GTG GAC TCC CTC TGG AGC GTG GCC      12046
Ala Ile Leu Leu Val Ser Ser Cys Val Asp Ser Leu Trp Ser Val Ala
4000                 4005                 4010                 4015

CAG GCC CTG TTG GTG CTG TGC CCT GGG ACT GGG CTC TCT ACC CTG TGT      12094
Gln Ala Leu Leu Val Leu Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys
                4020                 4025                 4030

CCT GCC GAG TCC TGG CAC CTG TCA CCC CTG CTG TGT GTG GGG CTC TGG      12142
Pro Ala Glu Ser Trp His Leu Ser Pro Leu Leu Cys Val Gly Leu Trp
            4035                 4040                 4045

GCA CTG CGG CTG TGG GGC GCC CTA CGG CTG GGG GCT GTT ATT CTC CGC      12190
Ala Leu Arg Leu Trp Gly Ala Leu Arg Leu Gly Ala Val Ile Leu Arg
        4050                 4055                 4060

TGG CGC TAC CAC GCC TTG CGT GGA GAG CTG TAC CGG CCG GCC TGG GAG      12238
Trp Arg Tyr His Ala Leu Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu
    4065                 4070                 4075
```

```
CCC CAG GAC TAC GAG ATG GTG GAG TTG TTC CTG CGC AGG CTG CGC CTC      12286
Pro Gln Asp Tyr Glu Met Val Glu Leu Phe Leu Arg Arg Leu Arg Leu
4080                4085                4090                4095

TGG ATG GGC CTC AGC AAG GTC AAG GAG TTC CGC CAC AAA GTC CGC TTT      12334
Trp Met Gly Leu Ser Lys Val Lys Glu Phe Arg His Lys Val Arg Phe
                    4100                4105                4110

GAA GGG ATG GAG CCG CTG CCC TCT CGC TCC TCC AGG GGC TCC AAG GTA      12382
Glu Gly Met Glu Pro Leu Pro Ser Arg Ser Ser Arg Gly Ser Lys Val
            4115                4120                4125

TCC CCG GAT GTG CCC CCA CCC AGC GCT GGC TCC GAT GCC TCG CAC CCC      12430
Ser Pro Asp Val Pro Pro Pro Ser Ala Gly Ser Asp Ala Ser His Pro
        4130                4135                4140

TCC ACC TCC TCC AGC CAG CTG GAT GGG CTG AGC GTG AGC CTG GGC CGG      12478
Ser Thr Ser Ser Ser Gln Leu Asp Gly Leu Ser Val Ser Leu Gly Arg
    4145                4150                4155

CTG GGG ACA AGG TGT GAG CCT GAG CCC TCC CGC CTC CAA GCC GTG TTC      12526
Leu Gly Thr Arg Cys Glu Pro Glu Pro Ser Arg Leu Gln Ala Val Phe
4160                4165                4170                4175

GAG GCC CTG CTC ACC CAG TTT GAC CGA CTC AAC CAG GCC ACA GAG GAC      12574
Glu Ala Leu Leu Thr Gln Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp
                    4180                4185                4190

GTC TAC CAG CTG GAG CAG CAG CTG CAC AGC CTG CAA GGC CGC AGG AGC      12622
Val Tyr Gln Leu Glu Gln Gln Leu His Ser Leu Gln Gly Arg Arg Ser
            4195                4200                4205

AGC CGG GCG CCC GCC GGA TCT TCC CGT GGC CCA TCC CCG GGC CTG CGG      12670
Ser Arg Ala Pro Ala Gly Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg
        4210                4215                4220

CCA GCA CTG CCC AGC CGC CTT GCC CGG GCC AGT CGG GGT GTG GAC CTG      12718
Pro Ala Leu Pro Ser Arg Leu Ala Arg Ala Ser Arg Gly Val Asp Leu
    4225                4230                4235

GCC ACT GGC CCC AGC AGG ACA CCT TCG GGC CAA GAA CAA GGT CCA CCC      12766
Ala Thr Gly Pro Ser Arg Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro
4240                4245                4250                4255

CAG CAG CAC TTA GTC CTC CTT CCT GGC GGG GGT GGG CCG TGG AGT CGG      12814
Gln Gln His Leu Val Leu Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg
                    4260                4265                4270

AGT GGA CAC CGC TCA GTA TTA CTT TCT GCC GCT GTC AAG GCC GAG GGC      12862
Ser Gly His Arg Ser Val Leu Leu Ser Ala Ala Val Lys Ala Glu Gly
            4275                4280                4285

CAG GCA GAA TGG CTG CAC GTA GGT TCC CCA GAG AGC AGG CAG GGG CAT      12910
Gln Ala Glu Trp Leu His Val Gly Ser Pro Glu Ser Arg Gln Gly His
        4290                4295                4300

CTG TCT GTC TGT GGG CTT CAG CAC TTT AAA GAG GCT GTG TGG CCA ACC      12958
Leu Ser Val Cys Gly Leu Gln His Phe Lys Glu Ala Val Trp Pro Thr
    4305                4310                4315
```

```
AGG ACC CAG GGT CCC CTC CCC AGC TCC CTT GGG AAG GAC ACA GCA GTA      13006
Arg Thr Gln Gly Pro Leu Pro Ser Ser Leu Gly Lys Asp Thr Ala Val
4320                4325                4330                4335

TTG GAC GGT TTC TAGCCTCTGA GATGCTAATT TATTTCCCCG AGTCCTCAGG          13058
Leu Asp Gly Phe

TACAGCGGGC TGTGCCCGGC CCCACCCCCT GGGCAGATGT CCCCCACTGC TAAGGCTGCT   13118

GGCTTCAGGG AGGGTTAGCC TGCACCGCCG CCACCCTGCC CCTAAGTTAT TACCTCTCCA   13178

GTTCCTACCG TACTCCCTGC ACCGTCTCAC TGTGTGTCTC GTGTCAGTAA TTTATATGGT   13238

GTTAAAATGT GTATATTTTT GTATGTCACT ATTTTCACTA GGGCTGAGGG GCCTGCGCCC   13298

AGAGCTGGCC TCCCCCAACA CCTGCTGCGC TTGGTAGGTG TGGTGGCGTT ATGGCAGCCC   13358

GGCTGCTGCT TGGATGCGAG CTTGGCCTTG GGCCGGTGCT GGGGGCACAG CTGTCTGCCA   13418

GGCACTCTCA TCACCCCAGA GGCCTTGTCA TCCTCCCTTG CCCCAGGCCA GGTAGCAAGA   13478

GAGCAGCGCC CAGGCCTGCT GGCATCAGGT CTGGGCAAGT AGCAGGACTA GGCATGTCAG   13538

AGGACCCCAG GGTGGTTAGA GGAAAAGACT CCTCCTGGGG GCTGGCTCCC AGGGTGGAGG   13598

AAGGTGACTG TGTGTGTGTG TGTGTGCGCG CGCGACGCGC GAGTGTGCTG TATGGCCCAG   13658

GCAGCCTCAA GGCCCTCGGA GCTGGCTGTG CCTGCTTCTG TGTACCACTT CTGTGGGCAT   13718

GGCCGCTTCT AGAGCCTCGA CACCCCCCCA ACCCCCGCAC CAAGCAGACA AAGTCAATAA   13778

AAGAGCTGTC TGACTGCAAA AAAAAAAAA                                      13807
```

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Gly Ala Ala Cys Arg Val Asn Cys Ser Gly Arg Gly Leu Arg Thr Leu
 1               5                  10                  15

Gly Pro Ala Leu Arg Ile Pro Ala Asp Ala Thr Ala Leu Asp Val Ser
            20                  25                  30

His Asn Leu Leu Arg Ala Leu Asp Val Gly Leu Leu Ala Asn Leu Ser
        35                  40                  45

Ala Leu Ala Glu Leu Asp Ile Ser Asn Asn Lys Ile Ser Thr Leu Glu
    50                  55                  60

Glu Gly Ile Phe Ala Asn Leu Phe Asn Leu Ser Glu Ile Asn Leu Ser
65                  70                  75                  80

Gly Asn Pro Phe Glu Cys Asp Cys Gly Leu Ala Trp Leu Pro Arg Trp
                85                  90                  95

Ala Glu Glu Gln Gln Val Arg Val Val Gln Pro Glu Ala Ala Thr Cys
            100                 105                 110

Ala Gly Pro Gly Ser Leu Ala Gly Gln Pro Leu Leu Gly Ile Pro Leu
        115                 120                 125
```

Leu Asp Ser Gly Cys Gly Glu Glu Tyr Val Ala Cys Leu Pro Asp Asn
130 135 140

Ser Ser Gly Thr Val Ala Ala Val Ser Phe Ser Ala Ala His Glu Gly
145 150 155 160

Leu Leu Gln Pro Glu Ala Cys Ser Ala Phe Cys Phe Ser Thr Gly Gln
165 170 175

Gly Leu Ala Ala Leu Ser Glu Gln Gly Trp Cys Leu Cys Gly Ala Ala
180 185 190

Gln Pro Ser Ser Ala Ser Phe Ala Cys Leu Ser Leu Cys Ser Gly Pro
195 200 205

Pro Pro Pro Pro Ala Pro Thr Cys Arg Gly Pro Thr Leu Leu Gln His
210 215 220

Val Phe Pro Ala Ser Pro Gly Ala Thr Leu Val Gly Pro His Gly Pro
225 230 235 240

Leu Ala Ser Gly Gln Leu Ala Ala Phe His Ile Ala Ala Pro Leu Pro
245 250 255

Val Thr Ala Thr Arg Trp Asp Phe Gly Asp Gly Ser Ala Glu Val Asp
260 265 270

Ala Ala Gly Pro Ala Ala Ser His Arg Tyr Val Leu Pro Gly Arg Tyr
275 280 285

His Val Thr Ala Val Leu Ala Leu Gly Ala Gly Ser Ala Leu Leu Gly
290 295 300

Thr Asp Val Gln Val Glu Ala Ala Pro Ala Ala Leu Glu Leu Val Cys
305 310 315 320

Pro Ser Ser Val Gln Ser Asp Glu Ser Leu Asp Leu Ser Ile Gln Asn
325 330 335

Arg Gly Gly Ser Gly Leu Glu Ala Ala Tyr Ser Ile Val Ala Leu Gly
340 345 350

Glu Glu Pro Ala Arg Ala Val His Pro Leu Cys Pro Ser Asp Thr Glu
355 360 365

Ile Phe Pro Gly Asn Gly His Cys Tyr Arg Leu Val Val Glu Lys Ala
370 375 380

Ala Trp Leu Gln Ala Gln Glu Gln Cys Gln Ala Trp Ala Gly Ala Ala
385 390 395 400

Leu Ala Met Val Asp Ser Pro Ala Val Gln Arg Phe Leu Val Ser Arg
405 410 415

Val Thr Arg Ser Leu Asp Val Trp Ile Gly Phe Ser Thr Val Gln Gly
420 425 430

Val Glu Val Gly Pro Ala Pro Gln Gly Glu Ala Phe Ser Leu Glu Ser
435 440 445

Cys Gln Asn Trp Leu Pro Gly Glu Pro His Pro Ala Thr Ala Glu His
450 455 460

Cys Val Arg Leu Gly Pro Thr Gly Trp Cys Asn Thr Asp Leu Cys Ser
465 470 475 480

Ala Pro His Ser Tyr Val Cys Glu Leu Gln Pro Gly Gly Pro Val Gln
485 490 495

Asp Ala Glu Asn Leu Leu Val Gly Ala Pro Ser Gly Asp Leu Gln Gly
500 505 510

Pro Leu Thr Pro Leu Ala Gln Gln Asp Gly Leu Ser Ala Pro His Glu
515 520 525

Pro Val Glu Val Met Val Phe Pro Gly Leu Arg Leu Ser Arg Glu Ala
530 535 540

Phe Leu Thr Thr Ala Glu Phe Gly Thr Gln Glu Leu Arg Arg Pro Ala
545 550 555 560

Gln Leu Arg Leu Gln Val Tyr Arg Leu Leu Ser Thr Ala Gly Thr Pro
565 570 575

Glu Asn Gly Ser Glu Pro Glu Ser Arg Ser Pro Asp Asn Arg Thr Gln
580 585 590

Leu Ala Pro Ala Cys Met Pro Gly Gly Arg Trp Cys Pro Gly Ala Asn
595 600 605

Ile Cys Leu Pro Leu Asp Ala Ser Cys His Pro Gln Ala Cys Ala Asn
610 615 620

Gly Cys Thr Ser Gly Pro Gly Leu Pro Gly Ala Pro Tyr Ala Leu Trp
625 630 635 640

Arg Glu Phe Leu Phe Ser Val Ala Ala Gly Pro Pro Ala Gln Tyr Ser
645 650 655

Val Thr Leu His Gly Gln Asp Val Leu Met Leu Pro Gly Asp Leu Val
660 665 670

Gly Leu Gln His Asp Ala Gly Pro Gly Ala Leu Leu His Cys Ser Pro
675 680 685

Ala Pro Gly His Pro Gly Pro Gln Ala Pro Tyr Leu Ser Ala Asn Ala
690 695 700

Ser Ser Trp Leu Pro His Leu Pro Ala Gln Leu Glu Gly Thr Trp Ala
705 710 715 720

Cys Pro Ala Cys Ala Leu Arg Leu Leu Ala Ala Thr Glu Gln Leu Thr
725 730 735

Val Leu Leu Gly Leu Arg Pro Asn Pro Gly Leu Arg Met Pro Gly Arg
740 745 750

Tyr Glu Val Arg Ala Glu Val Gly Asn Gly Val Ser Arg His Asn Leu
755 760 765

Ser Cys Ser Phe Asp Val Val Ser Pro Val Ala Gly Leu Arg Val Ile
770 775 780

Tyr Pro Ala Pro Arg Asp Gly Arg Leu Tyr Val Pro Thr Asn Gly Ser
785 790 795 800

Ala Leu Val Leu Gln Val Asp Ser Gly Ala Asn Ala Thr Ala Thr Ala
805 810 815

Arg Trp Pro Gly Gly Ser Val Ser Ala Arg Phe Glu Asn Val Cys Pro
820 825 830

Ala Leu Val Ala Thr Phe Val Pro Gly Cys Pro Trp Glu Thr Asn Asp
835 840 845

Thr Leu Phe Ser Val Val Ala Leu Pro Trp Leu Ser Glu Gly Glu His
850 855 860

Val Val Asp Val Val Val Glu Asn Ser Ala Ser Arg Ala Asn Leu Ser
865 870 875 880

Leu Arg Val Thr Ala Glu Glu Pro Ile Cys Gly Leu Arg Ala Thr Pro
885 890 895

Ser Pro Glu Ala Arg Val Leu Gln Gly Val Leu Val Arg Tyr Ser Pro
900 905 910

Val Val Glu Ala Gly Ser Asp Met Val Phe Arg Trp Thr Ile Asn Asp
915 920 925

Lys Gln Ser Leu Thr Phe Gln Asn Val Val Phe Asn Val Ile Tyr Gln
930 935 940

Ser Ala Ala Val Phe Lys Leu Ser Leu Thr Ala Ser Asn His Val Ser
945 950 955 960

Asn Val Thr Val Asn Tyr Asn Val Thr Val Glu Arg Met Asn Arg Met
965 970 975

Gln Gly Leu Gln Val Ser Thr Val Pro Ala Val Leu Ser Pro Asn Ala
980 985 990

Thr Leu Val Leu Thr Gly Gly Val Leu Val Asp Ser Ala Val Glu Val
995 1000 1005

Ala Phe Leu Trp Asn Phe Gly Asp Gly Glu Gln Ala Leu His Gln Phe
1010 1015 1020

Gln Pro Pro Tyr Asn Glu Ser Phe Pro Val Pro Asp Pro Ser Val Ala
1025 1030 1035 1040

Gln Val Leu Val Glu His Asn Val Met His Thr Tyr Ala Ala Pro Gly
1045 1050 1055

Glu Tyr Leu Leu Thr Val Leu Ala Ser Asn Ala Phe Glu Asn Leu Thr
1060 1065 1070

Gln Gln Val Pro Val Ser Val Arg Ala Ser Leu Pro Ser Val Ala Val
1075 1080 1085

Gly Val Ser Asp Gly Val Leu Val Ala Gly Arg Pro Val Thr Phe Tyr
1090 1095 1100

Pro His Pro Leu Pro Ser Pro Gly Gly Val Leu Tyr Thr Trp Asp Phe
1105 1110 1115 1120

Gly Asp Gly Ser Pro Val Leu Thr Gln Ser Gln Pro Ala Ala Asn His
1125 1130 1135

Thr Tyr Ala Ser Arg Gly Thr Tyr His Val Arg Leu Glu Val Asn Asn
1140 1145 1150

Thr Val Ser Gly Ala Ala Ala Gln Ala Asp Val Arg Val Phe Glu Glu
1155 1160 1165

Leu Arg Gly Leu Ser Val Asp Met Ser Leu Ala Val Glu Gln Gly Ala
1170 1175 1180

Pro Val Val Val Ser Ala Ala Val Gln Thr Gly Asp Asn Ile Thr Trp
1185 1190 1195 1200

Thr Phe Asp Met Gly Asp Gly Thr Val Leu Ser Gly Pro Glu Ala Thr
1205 1210 1215

Val Glu His Val Tyr Leu Arg Ala Gln Asn Cys Thr Val Thr Val Gly
1220 1225 1230

Ala Ala Ser Pro Ala Gly His Leu Ala Arg Ser Leu His Val Leu Val
1235 1240 1245

Phe Val Leu Glu Val Leu Arg Val Glu Pro Ala Ala Cys Ile Pro Thr
1250 1255 1260

Gln Pro Asp Ala Arg Leu Thr Ala Tyr Val Thr Gly Asn Pro Ala His
1265 1270 1275 1280

Tyr Leu Phe Asp Trp Thr Phe Gly Asp Gly Ser Ser Asn Thr Thr Val
1285 1290 1295

Arg Gly Cys Pro Thr Val Thr His Asn Phe Thr Arg Ser Gly Thr Phe
1300 1305 1310

Pro Leu Ala Leu Val Leu Ser Ser Arg Val Asn Arg Ala His Tyr Phe
1315 1320 1325

Thr Ser Ile Cys Val Glu Pro Glu Val Gly Asn Val Thr Leu Gln Pro
1330 1335 1340

Glu Arg Gln Phe Val Gln Leu Gly Asp Glu Ala Trp Leu Val Ala Cys
1345 1350 1355 1360

Ala Trp Pro Pro Phe Pro Tyr Arg Tyr Thr Trp Asp Phe Gly Thr Glu
1365 1370 1375

Glu Ala Ala Pro Thr Arg Ala Arg Gly Pro Glu Val Thr Phe Ile Tyr
1380 1385 1390

Arg Asp Pro Gly Ser Tyr Leu Val Thr Val Thr Ala Ser Asn Asn Ile
1395 1400 1405

Ser Ala Ala Asn Asp Ser Ala Leu Val Glu Val Gln Glu Pro Val Leu
1410 1415 1420

Val Thr Ser Ile Lys Val Asn Gly Ser Leu Gly Leu Glu Leu Gln Gln
1425 1430 1435 1440

Pro Tyr Leu Phe Ser Ala Val Gly Arg Gly Arg Pro Ala Ser Tyr Leu
1445 1450 1455

Trp Asp Leu Gly Asp Gly Gly Trp Leu Glu Gly Pro Glu Val Thr His
1460 1465 1470

Ala Tyr Asn Ser Thr Gly Asp Phe Thr Val Arg Val Ala Gly Trp Asn
1475 1480 1485

Glu Val Ser Arg Ser Glu Ala Trp Leu Asn Val Thr Val Lys Arg Arg
1490 1495 1500

Val Arg Gly Leu Val Val Asn Ala Ser Arg Thr Val Val Pro Leu Asn
1505 1510 1515 1520

Gly Ser Val Ser Phe Ser Thr Ser Leu Glu Ala Gly Ser Asp Val Arg
1525 1530 1535

Tyr Ser Trp Val Leu Cys Asp Arg Cys Thr Pro Ile Pro Gly Gly Pro
1540 1545 1550

Thr Ile Ser Tyr Thr Phe Arg Ser Val Gly Thr Phe Asn Ile Ile Val
1555 1560 1565

Thr Ala Glu Asn Glu Val Gly Ser Ala Gln Asp Ser Ile Phe Val Tyr
1570 1575 1580

Val Leu Gln Leu Ile Glu Gly Leu Gln Val Val Gly Gly Gly Arg Tyr
1585 1590 1595 1600

Phe Pro Thr Asn His Thr Val Gln Leu Gln Ala Val Val Arg Asp Gly
1605 1610 1615

Thr Asn Val Ser Tyr Ser Trp Thr Ala Trp Arg Asp Arg Gly Pro Ala
1620 1625 1630

Leu Ala Gly Ser Gly Lys Gly Phe Ser Leu Thr Val Leu Glu Ala Gly
1635 1640 1645

Thr Tyr His Val Gln Leu Arg Ala Thr Asn Met Leu Gly Ser Ala Trp
1650 1655 1660

Ala Asp Cys Thr Met Asp Phe Val Glu Pro Val Gly Trp Leu Met Val
1665 1670 1675 1680

Thr Ala Ser Pro Asn Pro Ala Ala Val Asn Thr Ser Val Thr Leu Ser
1685 1690 1695

Ala Glu Leu Ala Gly Gly Ser Gly Val Val Tyr Thr Trp Ser Leu Glu
1700 1705 1710

Glu Gly Leu Ser Trp Glu Thr Ser Glu Pro Phe Thr Thr His Ser Phe
1715 1720 1725

Pro Thr Pro Gly Leu His Leu Val Thr Met Thr Ala Gly Asn Pro Leu
1730 1735 1740

Gly Ser Ala Asn Ala Thr Val Glu Val Asp Val Gln Val Pro Val Ser
1745 1750 1755 1760

Gly Leu Ser Ile Arg Ala Ser Glu Pro Gly Gly Ser Phe Val Ala Ala
1765 1770 1775

Gly Ser Ser Val Pro Phe Trp Gly Gln Leu Ala Thr Gly Thr Asn Val
1780 1785 1790

Ser Trp Cys Trp Ala Val Pro Gly Gly Ser Ser Lys Arg Gly Pro His
1795 1800 1805

Val Thr Met Val Phe Pro Asp Ala Gly Thr Phe Ser Ile Arg Leu Asn
1810 1815 1820

Ala Ser Asn Ala Val Ser Trp Val Ser Ala Thr Tyr Asn Leu Thr Ala
1825 1830 1835 1840

Glu Glu Pro Ile Val Gly Leu Val Leu Trp Ala Ser Ser Lys Val Val
1845 1850 1855

Ala Pro Gly Gln Leu Val His Phe Gln Ile Leu Leu Ala Ala Gly Ser
1860 1865 1870

Ala Val Thr Phe Arg Leu Gln Val Gly Gly Ala Asn Pro Glu Val Leu
1875 1880 1885

Pro Gly Pro Arg Phe Ser His Ser Phe Pro Arg Val Gly Asp His Val
1890 1895 1900

Val Ser Val Arg Gly Lys Asn His Val Ser Trp Ala Gln Ala Gln Val
1905 1910 1915 1920

Arg Ile Val Val Leu Glu Ala Val Ser Gly Leu Gln Met Pro Asn Cys
1925 1930 1935

Cys Glu Pro Gly Ile Ala Thr Gly Thr Glu Arg Asn Phe Thr Ala Arg
1940 1945 1950

Val Gln Arg Gly Ser Arg Val Ala Tyr Ala Trp Tyr Phe Ser Leu Gln
1955 1960 1965

Lys Val Gln Gly Asp Ser Leu Val Ile Leu Ser Gly Arg Asp Val Thr
1970 1975 1980

Tyr Thr Pro Val Ala Ala Gly Leu Leu Glu Ile Gln Val Arg Ala Phe
1985 1990 1995 2000

Asn Ala Leu Gly Ser Glu Asn Arg Thr Leu Val Leu Glu Val Gln Asp
2005 2010 2015

Ala Val Gln Tyr Val Ala Leu Gln Ser Gly Pro Cys Phe Thr Asn Arg
2020 2025 2030

Ser Ala Gln Phe Glu Ala Ala Thr Ser Pro Ser Pro Arg Arg Val Ala
2035 2040 2045

Tyr His Trp Asp Phe Gly Asp Gly Ser Pro Gly Gln Asp Thr Asp Glu
2050 2055 2060

Pro Arg Ala Glu His Ser Tyr Leu Arg Pro Gly Asp Tyr Arg Val Gln
2065 2070 2075 2080

Val Asn Ala Ser Asn Leu Val Ser Phe Phe Val Ala Gln Ala Thr Val
2085 2090 2095

Thr Val Gln Val Leu Ala Cys Arg Glu Pro Glu Val Asp Val Val Leu
2100 2105 2110

Pro Leu Gln Val Leu Met Arg Arg Ser Gln Arg Asn Tyr Leu Glu Ala
2115 2120 2125

His Val Asp Leu Arg Asp Cys Val Thr Tyr Gln Thr Glu Tyr Arg Trp
2130 2135 2140

Glu Val Tyr Arg Thr Ala Ser Cys Gln Arg Pro Gly Arg Pro Ala Arg
2145 2150 2155 2160

Val Ala Leu Pro Gly Val Asp Val Ser Arg Pro Arg Leu Val Leu Pro
2165 2170 2175

Arg Leu Ala Leu Pro Val Gly His Tyr Cys Phe Val Phe Val Val Ser
2180 2185 2190

Phe Gly Asp Thr Pro Leu Thr Gln Ser Ile Gln Ala Asn Val Thr Val
2195 2200 2205

Ala Pro Glu Arg Leu Val Pro Ile Ile Glu Gly Gly Ser Tyr Arg Val
2210 2215 2220

Trp Ser Asp Thr Arg Asp Leu Val Leu Asp Gly Ser Glu Ser Tyr Asp
2225 2230 2235 2240

Pro Asn Leu Glu Asp Gly Asp Gln Thr Pro Leu Ser Phe His Trp Ala
2245 2250 2255

Cys Val Ala Ser Thr Gln Arg Glu Ala Gly Gly Cys Ala Leu Asn Phe
2260 2265 2270

Gly Pro Arg Gly Ser Ser Thr Val Thr Ile Pro Arg Glu Arg Leu Ala
2275 2280 2285

Ala Gly Val Glu Tyr Thr Phe Ser Leu Thr Val Trp Lys Ala Gly Arg
2290 2295 2300

Lys Glu Glu Ala Thr Asn Gln Thr Val Leu Ile Arg Ser Gly Arg Val
2305 2310 2315 2320

Pro Ile Val Ser Leu Glu Cys Val Ser Cys Lys Ala Gln Ala Val Tyr
2325 2330 2335

Glu Val Ser Arg Ser Ser Tyr Val Tyr Leu Glu Gly Arg Cys Leu Asn
2340 2345 2350

Cys Ser Ser Gly Ser Lys Arg Gly Arg Trp Ala Ala Arg Thr Phe Ser
2355 2360 2365

Asn Lys Thr Leu Val Leu Asp Glu Thr Thr Thr Ser Thr Gly Ser Ala
2370 2375 2380

Gly Met Arg Leu Val Leu Arg Arg Gly Val Leu Arg Asp Gly Glu Gly
2385 2390 2395 2400

Tyr Thr Phe Thr Leu Thr Val Leu Gly Arg Ser Gly Glu Glu Glu Gly
2405 2410 2415

Cys Ala Ser Ile Arg Leu Ser Pro Asn Arg Pro Pro Leu Gly Gly Ser
2420 2425 2430

Cys Arg Leu Phe Pro Leu Gly Ala Val His Ala Leu Thr Thr Lys Val
2435 2440 2445

His Phe Glu Cys Thr Gly Trp His Asp Ala Glu Asp Ala Gly Ala Pro
2450 2455 2460

Leu Val Tyr Ala Leu Leu Leu Arg Arg Cys Arg Gln Gly His Cys Glu
2465 2470 2475 2480

Glu Phe Cys Val Tyr Lys Gly Ser Leu Ser Ser Tyr Gly Ala Val Leu
2485 2490 2495

Pro Pro Gly Phe Arg Pro His Phe Glu Val Gly Leu Ala Val Val Val
2500 2505 2510

Gln Asp Gln Leu Gly Ala Ala Val Val Ala Leu Asn Arg Ser Leu Ala
2515 2520 2525

Ile Thr Leu Pro Glu Pro Asn Gly Ser Ala Thr Gly Leu Thr Val Trp
2530 2535 2540

Leu His Gly Leu Thr Ala Ser Val Leu Pro Gly Leu Leu Arg Gln Ala
2545 2550 2555 2560

Asp Pro Gln His Val Ile Glu Tyr Ser Leu Ala Leu Val Thr Val Leu
2565 2570 2575

Asn Glu Tyr Glu Arg Ala Leu Asp Val Ala Ala Glu Pro Lys His Glu
2580 2585 2590

Arg Gln His Arg Ala Gln Ile Arg Lys Asn Ile Thr Glu Thr Leu Val
2595 2600 2605

Ser Leu Arg Val His Thr Val Asp Asp Ile Gln Gln Ile Ala Ala Ala
2610 2615 2620

Leu Ala Gln Cys Met Gly Pro Ser Arg Glu Leu Val Cys Arg Ser Cys
2625 2630 2635 2640

Leu Lys Gln Thr Leu His Lys Leu Glu Ala Met Met Leu Ile Leu Gln
2645 2650 2655

Ala Glu Thr Thr Ala Gly Thr Val Thr Pro Thr Ala Ile Gly Asp Ser
2660 2665 2670

Ile Leu Asn Ile Thr Gly Asp Leu Ile His Leu Ala Ser Ser Asp Val
2675 2680 2685

Arg Ala Pro Gln Pro Ser Glu Leu Gly Ala Glu Ser Pro Ser Arg Met
2690 2695 2700

Val Ala Ser Gln Ala Tyr Asn Leu Thr Ser Ala Leu Met Arg Ile Leu
2705 2710 2715 2720

Met Arg Ser Arg Val Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu
2725 2730 2735

Glu Ile Val Ala Gln Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys
2740 2745 2750

Tyr Gly Gly Ala Pro Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala
2755 2760 2765

Phe Ser Gly Ala Leu Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe
2770 2775 2780

Leu Val Asp Ser Asn Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr
2785 2790 2795 2800

Val Ser Thr Lys Val Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala
2805 2810 2815

Gln Ile Pro Ile Glu Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys
2820 2825 2830

Val Pro Asn Asn Ser Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala
2835 2840 2845

Asn Ser Ala Asn Ser Val Val Val Gln Pro Gln Ala Ser Val Gly Ala
2850 2855 2860

Val Val Thr Leu Asp Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln
2865 2870 2875 2880

Leu Asn Tyr Thr Leu Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu
2885 2890 2895

Pro Tyr Leu Ala Val Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His
2900 2905 2910

Asn Cys Ser Ala Ser Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala
2915 2920 2925

Asp His Arg Pro Tyr Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro
2930 2935 2940

Ala Gly Ser Tyr His Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala
2945 2950 2955 2960

Leu Gln Val Ser Val Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser
2965 2970 2975

Glu Glu Asp Met Val Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu
2980 2985 2990

Thr Ser Pro Arg Gln Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe
2995 3000 3005

Gly Ala Ser Leu Phe Val Pro Pro Ser His Val Arg Phe Val Phe Pro
3010 3015 3020

Glu Pro Thr Ala Asp Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val
3025 3030 3035 3040

Cys Leu Val Thr Tyr Met Val Met Ala Ala Ile Leu His Lys Leu Asp
3045 3050 3055

Gln Leu Asp Ala Ser Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg
3060 3065 3070

Gly Arg Phe Lys Tyr Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly
3075 3080 3085

Ser Gly Thr Thr Ala His Val Gly Ile Met Leu Tyr Gly Val Asp Ser
3090 3095 3100

Arg Ser Gly His Arg His Leu Asp Gly Asp Arg Ala Phe His Arg Asn
3105 3110 3115 3120

Ser Leu Asp Ile Phe Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val
3125 3130 3135

Trp Lys Ile Arg Val Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp
3140 3145 3150

Phe Leu Gln His Val Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala
3155 3160 3165

Phe Phe Leu Val Asn Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly
3170 3175 3180

Gly Leu Val Glu Lys Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu
3185 3190 3195 3200

Arg Phe Arg Arg Leu Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp
3205 3210 3215

Lys His Ile Trp Leu Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe
3220 3225 3230

Thr Arg Ile Gln Arg Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe
3235 3240 3245

Leu Gly Ala Asn Ala Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr
3250 3255 3260

Ser Thr Gly His Val Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val
3265 3270 3275 3280

Ala Val Gly Leu Val Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala
3285 3290 3295

Ile Leu Phe Leu Phe Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro
3300 3305 3310

Ser Pro Thr Pro Ala Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu
3315 3320 3325

Asp Ser Ser Val Leu Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His
3330 3335 3340

Ala Glu Ala Phe Val Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp
3345 3350 3355 3360

Ser Lys Ser Leu Val Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp
3365 3370 3375

Pro Asp Leu Leu Ser Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln
3380 3385 3390

Leu Ala Arg Gly Gln Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly
3395 3400 3405

Phe Ser Leu Ala Ser Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser
3410 3415 3420

Asp Glu Asp Leu Ile Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro
3425 3430 3435 3440

Ala Pro Thr Gln Asp Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu
3445 3450 3455

Ser Ser Thr Pro Gly Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu
3460 3465 3470

Gly Glu Leu Gly Pro Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln
3475 3480 3485

Ala Ala Arg Leu Ser Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg
3490 3495 3500

Leu Leu Pro Ala Trp Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu
3505 3510 3515 3520

Leu Val Ala Val Ala Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe
3525 3530 3535

Pro Pro Gly Val Ser Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe
3540 3545 3550

Leu Ala Ser Phe Leu Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala
3555 3560 3565

Leu Tyr Phe Ser Leu Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp
3570 3575 3580

Thr Leu Val Glu Ser Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro
3585 3590 3595 3600

Arg Val Arg Pro Pro His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu
3605 3610 3615

Ala Arg Lys Val Lys Arg Leu His Gly Met Leu Arg Ser Leu Leu Val
3620 3625 3630

Tyr Met Leu Phe Leu Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala
3635 3640 3645

Ser Cys His Gly His Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu
3650 3655 3660

Leu His Ser Arg Ala Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp
3665 3670 3675 3680

Pro Trp Met Ala His Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser
3685 3690 3695

Ser Pro Glu Leu Gly Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu
3700 3705 3710

Ala Leu Tyr Pro Asp Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala
3715 3720 3725

Ala Gly Gly Phe Ser Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro
3730 3735 3740

His Asn Gly Ser Gly Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly
3745 3750 3755 3760

Ala Trp Ser Trp Gly Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val
3765 3770 3775

Gln Glu Leu Gly Leu Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe
3780 3785 3790

Leu Gln Leu His Asn Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu
3795 3800 3805

Glu Leu Thr Arg Tyr Ser Pro Ala Val Gly Leu His Ala Ala Val Thr
3810 3815 3820

Leu Arg Leu Glu Phe Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser
3825 3830 3835 3840

Val Arg Pro Phe Ala Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro
3845 3850 3855

Leu Leu Thr Ser Val Cys Leu Leu Leu Phe Ala Val His Phe Ala Val
3860 3865 3870

Ala Glu Ala Arg Thr Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg
3875 3880 3885

Leu Gly Ala Trp Ala Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr
3890 3895 3900

Ala Leu Val Arg Leu Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr
3905 3910 3915 3920

Arg Phe Val Arg Gly Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val
3925 3930 3935

Ala His Val Ser Ser Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe
3940 3945 3950

Leu Leu Leu Val Lys Ala Ala Gln His Val Arg Phe Val Arg Gln Trp
3955 3960 3965

Ser Val Phe Gly Lys Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly
3970 3975 3980

Val Thr Leu Gly Leu Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala
3985 3990 3995 4000

Ile Leu Leu Val Ser Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln
4005 4010 4015

Ala Leu Leu Val Leu Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro
4020 4025 4030

Ala Glu Ser Trp His Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala
4035 4040 4045

Leu Arg Leu Trp Gly Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp
4050 4055 4060

Arg Tyr His Ala Leu Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro
4065 4070 4075 4080

Gln Asp Tyr Glu Met Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp
4085 4090 4095

Met Gly Leu Ser Lys Val Lys Glu Phe Arg His Lys Val Arg Phe Glu
4100 4105 4110

Gly Met Glu Pro Leu Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser
4115 4120 4125

Pro Asp Val Pro Pro Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser
4130 4135 4140

Thr Ser Ser Ser Gln Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu
4145 4150 4155 4160

Gly Thr Arg Cys Glu Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu
4165 4170 4175

Ala Leu Leu Thr Gln Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val
4180 4185 4190

Tyr Gln Leu Glu Gln Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser
4195 4200 4205

Arg Ala Pro Ala Gly Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro
4210 4215 4220

Ala Leu Pro Ser Arg Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala
4225 4230 4235 4240

Thr Gly Pro Ser Arg Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln
4245 4250 4255

Gln His Leu Val Leu Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser
4260 4265 4270

Gly His Arg Ser Val Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln
4275 4280 4285

Ala Glu Trp Leu His Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu
4290 4295 4300

Ser Val Cys Gly Leu Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg
4305 4310 4315 4320

Thr Gln Gly Pro Leu Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu
4325 4330 4335

Asp Gly Phe

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3: (Compare Figure 7)

```
CTC AAC GAG GAG CCC CTG ACG CTG GCG GGC GAG GAG ATC GTG GCC CAG          48
Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu Glu Ile Val Ala Gln
4340                 4345                4350                4355

GGC AAG CGC TCG GAC CCG CGG AGC CTG CTG TGC TAT GGC GGC GCC CCA          96
Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys Tyr Gly Gly Ala Pro
                4360                4365                4370

GGG CCT GGC TGC CAC TTC TCC ATC CCC GAG GCT TTC AGC GGG GCC CTG         144
Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala Phe Ser Gly Ala Leu
            4375                4380                4385

GCC AAC CTC AGT GAC GTG GTG CAG CTC ATC TTT CTG GTG GAC TCC AAT         192
Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe Leu Val Asp Ser Asn
        4390                4395                4400

CCC TTT CCC TTT GGC TAT ATC AGC AAC TAC ACC GTC TCC ACC AAG GTG         240
Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr Val Ser Thr Lys Val
    4405                4410                4415

GCC TCG ATG GCA TTC CAG ACA CAG GCC GGC GCC CAG ATC CCC ATC GAG         288
Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala Gln Ile Pro Ile Glu
4420                4425                4430                4435

CGG CTG GCC TCA GAG CGC GCC ATC ACC GTG AAG GTG CCC AAC AAC TCG         336
Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys Val Pro Asn Asn Ser
                4440                4445                4450

GAC TGG GCT GCC CGG GGC CAC CGC AGC TCC GCC AAC TCC GCC AAC TCC         384
Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala Asn Ser Ala Asn Ser
            4455                4460                4465

GTT GTG GTC CAG CCC CAG GCC TCC GTC GGT GCT GTG GTC ACC CTG GAC         432
Val Val Val Gln Pro Gln Ala Ser Val Gly Ala Val Val Thr Leu Asp
        4470                4475                4480

AGC AGC AAC CCT GCG GCC GGG CTG CAT CTG CAG CTC AAC TAT ACG CTG         480
Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln Leu Asn Tyr Thr Leu
    4485                4490                4495

CTG GAC GGC CAC TAC CTG TCT GAG GAA CCT GAG CCC TAC CTG GCA GTC         528
Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu Pro Tyr Leu Ala Val
4500                4505                4510                4515

TAC CTA CAC TCG GAG CCC CGG CCC AAT GAG CAC AAC TGC TCG GCT AGC         576
Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His Asn Cys Ser Ala Ser
                4520                4525                4530

AGG AGG ATC CGC CCA GAG TCA CTC CAG GGT GCT GAC CAC CGG CCC TAC         624
Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala Asp His Arg Pro Tyr
            4535                4540                4545

ACC TTC TTC ATT TCC CCG GGG AGC AGA GAC CCA GCG GGG AGT TAC CAT         672
Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro Ala Gly Ser Tyr His
        4550                4555                4560

CTG AAC CTC TCC AGC CAC TTC CGC TGG TCG GCG CTG CAG GTG TCC GTG         720
Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala Leu Gln Val Ser Val
    4565                4570                4575
```

```
GGC CTG TAC ACG TCC CTG TGC CAG TAC TTC AGC GAG GAG GAC ATG GTG          768
Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser Glu Glu Asp Met Val
4580                 4585                 4590                 4595

TGG CGG ACA GAG GGG CTG CTG CCC CTG GAG GAG ACC TCG CCC CGC CAG          816
Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu Thr Ser Pro Arg Gln
                4600                 4605                 4610

GCC GTC TGC CTC ACC CGC CAC CTC ACC GCC TTC GGC GCC AGC CTC TTC          864
Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe Gly Ala Ser Leu Phe
            4615                 4620                 4625

GTG CCC CCA AGC CAT GTC CGC TTT GTG TTT CCT GAG CCG ACA GCG GAT          912
Val Pro Pro Ser His Val Arg Phe Val Phe Pro Glu Pro Thr Ala Asp
        4630                 4635                 4640

GTA AAC TAC ATC GTC ATG CTG ACA TGT GCT GTG TGC CTG GTG ACC TAC          960
Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val Cys Leu Val Thr Tyr
    4645                 4650                 4655

ATG GTC ATG GCC GCC ATC CTG CAC AAG CTG GAC CAG TTG GAT GCC AGC         1008
Met Val Met Ala Ala Ile Leu His Lys Leu Asp Gln Leu Asp Ala Ser
4660                 4665                 4670                 4675

CGG GGC CGC GCC ATC CCT TTC TGT GGG CAG CGG GGC CGC TTC AAG TAC         1056
Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg Gly Arg Phe Lys Tyr
                4680                 4685                 4690

GAG ATC CTC GTC AAG ACA GGC TGG GGC CGG GGC TCA GGT ACC ACG GCC         1104
Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly Ser Gly Thr Thr Ala
            4695                 4700                 4705

CAC GTG GGC ATC ATG CTG TAT GGG GTG GAC AGC CGG AGC GGC CAC CGG         1152
His Val Gly Ile Met Leu Tyr Gly Val Asp Ser Arg Ser Gly His Arg
        4710                 4715                 4720

CAC CTG GAC GGC GAC AGA GCC TTC CAC CGC AAC AGC CTG GAC ATC TTC         1200
His Leu Asp Gly Asp Arg Ala Phe His Arg Asn Ser Leu Asp Ile Phe
    4725                 4730                 4735

CGG ATC GCC ACC CCG CAC AGC CTG GGT AGC GTG TGG AAG ATC CGA GTG         1248
Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val Trp Lys Ile Arg Val
4740                 4745                 4750                 4755

TGG CAC GAC AAC AAA GGG CTC AGC CCT GCC TGG TTC CTG CAG CAC GTC         1296
Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp Phe Leu Gln His Val
                4760                 4765                 4770

ATC GTC AGG GAC CTG CAG ACG GCA CGC AGC GCC TTC TTC CTG GTC AAT         1344
Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala Phe Phe Leu Val Asn
            4775                 4780                 4785

GAC TGG CTT TCG GTG GAG ACG GAG GCC AAC GGG GGC CTG GTG GAG AAG         1392
Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly Gly Leu Val Glu Lys
        4790                 4795                 4800

GAG GTG CTG GCC GCG AGC GAC GCA GCC CTT TTG CGC TTC CGG CGC CTG         1440
Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu Arg Phe Arg Arg Leu
    4805                 4810                 4815

CTG GTG GCT GAG CTG CAG CGT GGC TTC TTT GAC AAG CAC ATC TGG CTC         1488
Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp Lys His Ile Trp Leu
4820                 4825                 4830                 4835
```

```
TCC ATA TGG GAC CGG CCG CCT CGT AGC CGT TTC ACT CGC ATC CAG AGG      1536
Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe Thr Arg Ile Gln Arg
                4840                4845                4850

GCC ACC TGC TGC GTT CTC CTC ATC TGC CTC TTC CTG GGC GCC AAC GCC      1584
Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe Leu Gly Ala Asn Ala
            4855                4860                4865

GTG TGG TAC GGG GCT GTT GGC GAC TCT GCC TAC AGC ACG GGG CAT GTG      1632
Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr Ser Thr Gly His Val
        4870                4875                4880

TCC AGG CTG AGC CCG CTG AGC GTC GAC ACA GTC GCT GTT GGC CTG GTG      1680
Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val Ala Val Gly Leu Val
    4885                4890                4895

TCC AGC GTG GTT GTC TAT CCC GTC TAC CTG GCC ATC CTT TTT CTC TTC      1728
Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala Ile Leu Phe Leu Phe
4900                4905                4910                4915

CGG ATG TCC CGG AGC AAG GTG GCT GGG AGC CCG AGC CCC ACA CCT GCC      1776
Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro Ser Pro Thr Pro Ala
                4920                4925                4930

GGG CAG CAG GTG CTG GAC ATC GAC AGC TGC CTG GAC TCG TCC GTG CTG      1824
Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu Asp Ser Ser Val Leu
            4935                4940                4945

GAC AGC TCC TTC CTC ACG TTC TCA GGC CTC CAC GCT GAG GCC TTT GTT      1872
Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His Ala Glu Ala Phe Val
        4950                4955                4960

GGA CAG ATG AAG AGT GAC TTG TTT CTG GAT GAT TCT AAG AGT CTG GTG      1920
Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp Ser Lys Ser Leu Val
    4965                4970                4975

TGC TGG CCC TCC GGC GAG GGA ACG CTC AGT TGG CCG GAC CTG CTC AGT      1968
Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp Pro Asp Leu Leu Ser
4980                4985                4990                4995

GAC CCG TCC ATT GTG GGT AGC AAT CTG CGG CAG CTG GCA CGG GGC CAG      2016
Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln Leu Ala Arg Gly Gln
                5000                5005                5010

GCG GGC CAT GGG CTG GGC CCA GAG GAG GAC GGC TTC TCC CTG GCC AGC      2064
Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly Phe Ser Leu Ala Ser
            5015                5020                5025

CCC TAC TCG CCT GCC AAA TCC TTC TCA GCA TCA GAT GAA GAC CTG ATC      2112
Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser Asp Glu Asp Leu Ile
        5030                5035                5040

CAG CAG GTC CTT GCC GAG GGG GTC AGC AGC CCA GCC CCT ACC CAA GAC      2160
Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro Ala Pro Thr Gln Asp
    5045                5050                5055

2ACC CAC ATG GAA ACG GAC CTG CTC AGC AGC CTG TCC AGC ACT CCT GGG      2208
Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu Ser Ser Thr Pro Gly
5060                5065                5070                5075
```

GAG AAG ACA GAG ACG CTG GCG CTG CAG AGG CTG GGG GAG CTG GGG CCA 2256
Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu Gly Glu Leu Gly Pro
5080 5085 5090

8CCC AGC CCA GGC CTG AAC TGG GAA CAG CCC CAG GCA GCG AGG CTG TCC 2304
Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln Ala Ala Arg Leu Ser
5095 5100 5105

AGG ACA GGA CTG GTG GAG GGT CTG CGG AAG CGC CTG CTG CCG GCC TGG 2352
Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg Leu Leu Pro Ala Trp
5110 5115 5120

TGT GCC TCC CTG GCC CAC GGG CTC AGC CTG CTC CTG GTG GCT GTG GCT 2400
Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu Leu Val Ala Val Ala
5125 5130 5135

GTG GCT GTC TCA GGG TGG GTG GGT GCG AGC TTC CCC CCG GGC GTG AGT 2448
Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe Pro Pro Gly Val Ser
5140 5145 5150 5155

GTT GCG TGG CTC CTG TCC AGC AGC GCC AGC TTC CTG GCC TCA TTC CTC 2496
Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe Leu Ala Ser Phe Leu
5160 5165 5170

GGC TGG GAG CCA CTG AAG GTC TTG CTG GAA GCC CTG TAC TTC TCA CTG 2544
Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala Leu Tyr Phe Ser Leu
5175 5180 5185

GTG GCC AAG CGG CTG CAC CCG GAT GAA GAT GAC ACC CTG GTA GAG AGC 2592
Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp Thr Leu Val Glu Ser
5190 5195 5200

CCG GCT GTG ACG CCT GTG AGC GCA CGT GTG CCC CGC GTA CGG CCA CCC 2640
Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro Arg Val Arg Pro Pro
5205 5210 5215

CAC GGC TTT GCA CTC TTC CTG GCC AAG GAA GAA GCC CGC AAG GTC AAG 2688
His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu Ala Arg Lys Val Lys
5220 5225 5230 5235

AGG CTA CAT GGC ATG CTG CGG AGC CTC CTG GTG TAC ATG CTT TTT CTG 2736
Arg Leu His Gly Met Leu Arg Ser Leu Leu Val Tyr Met Leu Phe Leu
5240 5245 5250

CTG GTG ACC CTG CTG GCC AGC TAT GGG GAT GCC TCA TGC CAT GGG CAC 2784
Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala Ser Cys His Gly His
5255 5260 5265

GCC TAC CGT CTG CAA AGC GCC ATC AAG CAG GAG CTG CAC AGC CGG GCC 2832
Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu Leu His Ser Arg Ala
5270 5275 5280

TTC CTG GCC ATC ACG CGG TCT GAG GAG CTC TGG CCA TGG ATG GCC CAC 2880
Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp Pro Trp Met Ala His
5285 5290 5295

GTG CTG CTG CCC TAC GTC CAC GGG AAC CAG TCC AGC CCA GAG CTG GGG 2928
Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser Ser Pro Glu Leu Gly
5300 5305 5310 5315

CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG GAA GCA CTC TAC CCA GAC 2976

```
CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG GAA GCA CTC TAC CCA GAC      2976
Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu Ala Leu Tyr Pro Asp
                5320                5325                5330

CCT CCC GGC CCC AGG GTC CAC ACG TGC TCG GCC GCA GGA GGC TTC AGC      3024
Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala Ala Gly Gly Phe Ser
            5335                5340                5345

ACC AGC GAT TAC GAC GTT GGC TGG GAG AGT CCT CAC AAT GGC TCG GGG      3072
Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro His Asn Gly Ser Gly
        5350                5355                5360

ACG TGG GCC TAT TCA GCG CCG GAT CTG CTG GGG GCA TGG TCC TGG GGC      3120
Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly Ala Trp Ser Trp Gly
    5365                5370                5375

TCC TGT GCC GTG TAT GAC AGC GGG GGC TAC GTG CAG GAG CTG GGC CTG      3168
Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val Gln Glu Leu Gly Leu
5380                5385                5390                5395

AGC CTG GAG GAG AGC CGC GAC CGG CTG CGC TTC CTG CAG CTG CAC AAC      3216
Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe Leu Gln Leu His Asn
                5400                5405                5410

TGG CTG GAC AAC AGG AGC CGC GCT GTG TTC CTG GAG CTC ACG CGC TAC      3264
Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu Glu Leu Thr Arg Tyr
            5415                5420                5425

AGC CCG GCC GTG GGG CTG CAC GCC GCC GTC ACG CTG CGC CTC GAG TTC      3312
Ser Pro Ala Val Gly Leu His Ala Ala Val Thr Leu Arg Leu Glu Phe
        5430                5435                5440

CCG GCG GCC GGC CGC GCC CTG GCC GCC CTC AGC GTC CGC CCC TTT GCG      3360
Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser Val Arg Pro Phe Ala
    5445                5450                5455

CTG CGC CGC CTC AGC GCG GGC CTC TCG CTG CCT CTG CTC ACC TCG GTG      3408
Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro Leu Leu Thr Ser Val
5460                5465                5470                5475

TGC CTG CTG CTG TTC GCC GTG CAC TTC GCC GTG GCC GAG GCC CGT ACT      3456
Cys Leu Leu Leu Phe Ala Val His Phe Ala Val Ala Glu Ala Arg Thr
                5480                5485                5490

TGG CAC AGG GAA GGG CGC TGG CGC GTG CTG CGG CTC GGA GCC TGG GCG      3504
Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg Leu Gly Ala Trp Ala
            5495                5500                5505

CGG TGG CTG CTG GTG GCG CTG ACG GCG GCC ACG GCA CTG GTA CGC CTC      3552
Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr Ala Leu Val Arg Leu
        5510                5515                5520

GCC CAG CTG GGT GCC GCT GAC CGC CAG TGG ACC CGT TTC GTG CGC GGC      3600
Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr Arg Phe Val Arg Gly
    5525                5530                5535

CGC CCG CGC CGC TTC ACT AGC TTC GAC CAG GTG GCG CAC GTG AGC TCC      3648
Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val Ala His Val Ser Ser
5540                5545                5550                5555
```

GCA GCC CGT GGC CTG GCG GCC TCG CTG CTC TTC CTG CTT TTG GTC AAG 3696
Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe Leu Leu Leu Val Lys
5560 5565 5570

2GCT GCC CAG CAC GTA CGC TTC GTG CGC CAG TGG TCC GTC TTT GGC AAG 3744
Ala Ala Gln His Val Arg Phe Val Arg Gln Trp Ser Val Phe Gly Lys
5575 5580 5585

ACA TTA TGC CGA GCT CTG CCA GAG CTC CTG GGG GTC ACC TTG GGC CTG 3792
Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly Val Thr Leu Gly Leu
5590 5595 5600

GTG GTG CTC GGG GTA GCC TAC GCC CAG CTG GCC ATC CTG CTC GTG TCT 3840
Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala Ile Leu Leu Val Ser
5605 5610 5615

TCC TGT GTG GAC TCC CTC TGG AGC GTG GCC CAG GCC CTG TTG GTG CTG 3888
Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln Ala Leu Leu Val Leu
5620 5625 5630 5635

TGC CCT GGG ACT GGG CTC TCT ACC CTG TGT CCT GCC GAG TCC TGG CAC 3936
Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro Ala Glu Ser Trp His
5640 5645 5650

CTG TCA CCC CTG CTG TGT GTG GGG CTC TGG GCA CTG CGG CTG TGG GGC 3984
Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala Leu Arg Leu Trp Gly
5655 5660 5665

GCC CTA CGG CTG GGG GCT GTT ATT CTC CGC TGG CGC TAC CAC GCC TTG 4032
Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp Arg Tyr His Ala Leu
5670 5675 5680

CGT GGA GAG CTG TAC CGG CCG GCC TGG GAG CCC CAG GAC TAC GAG ATG 4080
Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro Gln Asp Tyr Glu Met
5685 5690 5695

GTG GAG TTG TTC CTG CGC AGG CTG CGC CTC TGG ATG GGC CTC AGC AAG 4128
Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp Met Gly Leu Ser Lys
5700 5705 5710 5715

GTC AAG GAG TTC CGC CAC AAA GTC CGC TTT GAA GGG ATG GAG CCG CTG 4176
Val Lys Glu Phe Arg His Lys Val Arg Phe Glu Gly Met Glu Pro Leu
5720 5725 5730

CCC TCT CGC TCC TCC AGG GGC TCC AAG GTA TCC CCG GAT GTG CCC CCA 4224
Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser Pro Asp Val Pro Pro
5735 5740 5745

CCC AGC GCT GGC TCC GAT GCC TCG CAC CCC TCC ACC TCC TCC AGC CAG 4272
Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser Thr Ser Ser Ser Gln
5750 5755 5760

CTG GAT GGG CTG AGC GTG AGC CTG GGC CGG CTG GGG ACA AGG TGT GAG 4320
Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu Gly Thr Arg Cys Glu
5765 5770 5775

CCT GAG CCC TCC CGC CTC CAA GCC GTG TTC GAG GCC CTG CTC ACC CAG 4368
Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu Ala Leu Leu Thr Gln
5780 5785 5790 5795

```
TTT GAC CGA CTC AAC CAG GCC ACA GAG GAC GTC TAC CAG CTG GAG CAG      4416
Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val Tyr Gln Leu Glu Gln
                5800                5805                5810

CAG CTG CAC AGC CTG CAA GGC CGC AGG AGC AGC CGG GCG CCC GCC GGA      4464
Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser Arg Ala Pro Ala Gly
            5815                5820                5825

TCT TCC CGT GGC CCA TCC CCG GGC CTG CGG CCA GCA CTG CCC AGC CGC      4512
Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro Ala Leu Pro Ser Arg
        5830                5835                5840

CTT GCC CGG GCC AGT CGG GGT GTG GAC CTG GCC ACT GGC CCC AGC AGG      4560
Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala Thr Gly Pro Ser Arg
    5845                5850                5855

ACA CCT TCG GGC CAA GAA CAA GGT CCA CCC CAG CAG CAC TTA GTC CTC      4608
Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln Gln His Leu Val Leu
5860                5865                5870                5875

CTT CCT GGC GGG GGT GGG CCG TGG AGT CGG AGT GGA CAC CGC TCA GTA      4656
Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser Gly His Arg Ser Val
                5880                5885                5890

TTA CTT TCT GCC GCT GTC AAG GCC GAG GGC CAG GCA GAA TGG CTG CAC      4704
Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln Ala Glu Trp Leu His
            5895                5900                5905

GTA GGT TCC CCA GAG AGC AGG CAG GGG CAT CTG TCT GTC TGT GGG CTT      4752
Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu Ser Val Cys Gly Leu
        5910                5915                5920

CAG CAC TTT AAA GAG GCT GTG TGG CCA ACC AGG ACC CAG GGT CCC CTC      4800
Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg Thr Gln Gly Pro Leu
    5925                5930                5935

CCC AGC TCC CTT GGG AAG GAC ACA GCA GTA TTG GAC GGT TTC              4842
Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu Asp Gly Phe
5940                5945                5950

TAGCCTCTGA GATGCTAATT TATTTCCCCG AGTCCTCAGG TACAGCGGGC TGTGCCCGGC    4902

CCCACCCCCT GGGCAGATGT CCCCCACTGC TAAGGCTGCT GGCTTCAGGG AGGGTTAGCC    4962

2TGCACCGCCG CCACCCTGCC CCTAAGTTAT TACCTCTCCA GTTCCTACCG TACTCCCTGC    5022

ACCGTCTCAC TGTGTGTCTC GTGTCAGTAA TTTATATGGT GTTAAAATGT GTATATTTTT    5082

GTATGTCACT ATTTTCACTA GGGCTGAGGG GCCTGCGCCC AGAGCTGGCC TCCCCCAACA    5142

CCTGCTGCGC TTGGTAGGTG TGGTGGCGTT ATGGCAGCCC GGCTGCTGCT TGGATGCGAG    5202

CTTGGCCTTG GGCCGGTGCT GGGGGCACAG CTGTCTGCCA GGCACTCTCA TCACCCCAGA    5262

GGCCTTGTCA TCCTCCCTTG CCCCAGGCCA GGTAGCAAGA GAGCAGCGCC CAGGCCTGCT    5322

GGCATCAGGT CTGGGCAAGT AGCAGGACTA GGCATGTCAG AGGACCCCAG GGTGGTTAGA    5382

GGAAAAGACT CCTCCTGGGG GCTGGCTCCC AGGGTGGAGG AAGGTGACTG TGTGTGTGTG    5442

TGTGTGCGCG CGCGACGCGC GAGTGTGCTG TATGGCCCAG GCAGCCTCAA GCCCCTCGGA    5502
```

GCTGGCTGTG CCTGCTTCTG TGTACCACTT CTGTGGGCAT GGCCGCTTCT AGAGCCTCGA 5562

CACCCCCCCA ACCCCCGCAC CAAGCAGACA AAGTCAATAA AAGAGCTGTC TGACTGCAAA 5622

AAAAAAAAA 5631

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: (Compare Figure 7)

Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu Glu Ile Val Ala Gln
1 5 10 15

Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys Tyr Gly Gly Ala Pro
20 25 30

Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala Phe Ser Gly Ala Leu
35 40 45

Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe Leu Val Asp Ser Asn
50 55 60

Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr Val Ser Thr Lys Val
65 70 75 80

Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala Gln Ile Pro Ile Glu
85 90 95

Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys Val Pro Asn Asn Ser
100 105 110

Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala Asn Ser Ala Asn Ser
115 120 125

Val Val Val Gln Pro Gln Ala Ser Val Gly Ala Val Val Thr Leu Asp
130 135 140

Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln Leu Asn Tyr Thr Leu
145 150 155 160

Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu Pro Tyr Leu Ala Val
165 170 175

Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His Asn Cys Ser Ala Ser
180 185 190

Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala Asp His Arg Pro Tyr
195 200 205

Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro Ala Gly Ser Tyr His
210 215 220

Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala Leu Gln Val Ser Val
225 230 235 240

Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser Glu Glu Asp Met Val
245 250 255

Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu Thr Ser Pro Arg Gln
260 265 270

Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe Gly Ala Ser Leu Phe
275 280 285

Val Pro Pro Ser His Val Arg Phe Val Phe Pro Glu Pro Thr Ala Asp
290 295 300

Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val Cys Leu Val Thr Tyr
305 310 315 320

Met Val Met Ala Ala Ile Leu His Lys Leu Asp Gln Leu Asp Ala Ser
325 330 335

Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg Gly Arg Phe Lys Tyr
340 345 350

Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly Ser Gly Thr Thr Ala
355 360 365

His Val Gly Ile Met Leu Tyr Gly Val Asp Ser Arg Ser Gly His Arg
370 375 380

His Leu Asp Gly Asp Arg Ala Phe His Arg Asn Ser Leu Asp Ile Phe
385 390 395 400

Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val Trp Lys Ile Arg Val
405 410 415

Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp Phe Leu Gln His Val
420 425 430

Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala Phe Phe Leu Val Asn
435 440 445

Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly Gly Leu Val Glu Lys
450 455 460

Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu Arg Phe Arg Arg Leu
465 470 475 480

Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp Lys His Ile Trp Leu
485 490 495

Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe Thr Arg Ile Gln Arg
500 505 510

Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe Leu Gly Ala Asn Ala
515 520 525

Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr Ser Thr Gly His Val
530 535 540

Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val Ala Val Gly Leu Val
545 550 555 560

Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala Ile Leu Phe Leu Phe
565 570 575

Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro Ser Pro Thr Pro Ala
580 585 590

Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu Asp Ser Ser Val Leu
595 600 605

Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His Ala Glu Ala Phe Val
610 615 620

Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp Ser Lys Ser Leu Val
625 630 635 640

Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp Pro Asp Leu Leu Ser
645 650 655

Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln Leu Ala Arg Gly Gln
660 665 670

Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly Phe Ser Leu Ala Ser
675 680 685

Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser Asp Glu Asp Leu Ile
690 695 700

Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro Ala Pro Thr Gln Asp
705 710 715 720

Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu Ser Ser Thr Pro Gly
725 730 735

Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu Gly Glu Leu Gly Pro
740 745 750

Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln Ala Ala Arg Leu Ser
755 760 765

Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg Leu Leu Pro Ala Trp
770 775 780

Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu Leu Val Ala Val Ala
785 790 795 800

Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe Pro Pro Gly Val Ser
805 810 815

Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe Leu Ala Ser Phe Leu
820 825 830

Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala Leu Tyr Phe Ser Leu
835 840 845

Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp Thr Leu Val Glu Ser
850 855 860

Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro Arg Val Arg Pro Pro
865 870 875 880

His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu Ala Arg Lys Val Lys
885 890 895

Arg Leu His Gly Met Leu Arg Ser Leu Leu Val Tyr Met Leu Phe Leu
900 905 910

Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala Ser Cys His Gly His
915 920 925

Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu Leu His Ser Arg Ala
930 935 940

Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp Pro Trp Met Ala His
945 950 955 960

Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser Ser Pro Glu Leu Gly
965 970 975

Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu Ala Leu Tyr Pro Asp
980 985 990

Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala Ala Gly Gly Phe Ser
995 1000 1005

Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro His Asn Gly Ser Gly
1010 1015 1020

Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly Ala Trp Ser Trp Gly
1025 1030 1035 1040

Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val Gln Glu Leu Gly Leu
1045 1050 1055

2
Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe Leu Gln Leu His Asn
1060 1065 1070

Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu Glu Leu Thr Arg Tyr
1075 1080 1085

Ser Pro Ala Val Gly Leu His Ala Ala Val Thr Leu Arg Leu Glu Phe
1090 1095 1100

Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser Val Arg Pro Phe Ala
1105 1110 1115 1120

Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro Leu Leu Thr Ser Val
1125 1130 1135

Cys Leu Leu Leu Phe Ala Val His Phe Ala Val Ala Glu Ala Arg Thr
1140 1145 1150

Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg Leu Gly Ala Trp Ala
1155 1160 1165

Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr Ala Leu Val Arg Leu
1170 1175 1180

8
Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr Arg Phe Val Arg Gly
1185 1190 1195 1200

Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val Ala His Val Ser Ser
2 1205 1210 1215

Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe Leu Leu Leu Val Lys
1220 1225 1230

Ala Ala Gln His Val Arg Phe Val Arg Gln Trp Ser Val Phe Gly Lys
1235 1240 1245

Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly Val Thr Leu Gly Leu
1250 1255 1260

Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala Ile Leu Leu Val Ser
1265 1270 1275 1280

Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln Ala Leu Leu Val Leu
1285 1290 1295

Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro Ala Glu Ser Trp His
1300 1305 1310

Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala Leu Arg Leu Trp Gly
1315 1320 1325

Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp Arg Tyr His Ala Leu
1330 1335 1340

Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro Gln Asp Tyr Glu Met
1345 1350 1355 1360

Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp Met Gly Leu Ser Lys
1365 1370 1375

Val Lys Glu Phe Arg His Lys Val Arg Phe Glu Gly Met Glu Pro Leu
1380 1385 1390

Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser Pro Asp Val Pro Pro
1395 1400 1405

Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser Thr Ser Ser Ser Gln
1410 1415 1420

Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu Gly Thr Arg Cys Glu
1425 1430 1435 1440

Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu Ala Leu Leu Thr Gln
1445 1450 1455

Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val Tyr Gln Leu Glu Gln
1460 1465 1470

Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser Arg Ala Pro Ala Gly
1475 1480 1485

Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro Ala Leu Pro Ser Arg
1490 1495 1500

Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala Thr Gly Pro Ser Arg
1505 1510 1515 1520

Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln Gln His Leu Val Leu
1525 1530 1535

Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser Gly His Arg Ser Val
1540 1545 1550

Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln Ala Glu Trp Leu His
1555 1560 1565

Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu Ser Val Cys Gly Leu
1570 1575 1580

Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg Thr Gln Gly Pro Leu
1585 1590 1595 1600

Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu Asp Gly Phe
1605 1610

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5: (Compare Figure 8)

AGCTTGGCAC CATCAAGGGC CAGTTCAACT TTGTCCACGT GATCGTCACC CCGCTGGACT 60

ACGAGTGCAA CCTGGTGTCC CTGCAGTGCA GGAAAGACAT GGAGGGCCTT GTGGACACCA 120

GCGTGGCCAA GATCGTGTCT GACCGCAACC TGCCCTTCGT GGCCCCGCCAG ATGGCCCTGC 180

ACGCAAATAT GGCCTCACAG GTGCATCATA GCCGCTCCAA CCCCACCGAT ATCTACCCCT 240

CCAAGTGGAT TGCCCCGCTC CGCCACATCA AGCGGCTCCG CCAGCGGATC TGCGAGGAAG 300

CCGCCTACTC CAACCCCAGC CTACCTCTGG TGCACCCTCC GTCCCATAGC AAAGCCCCTG 360

CACAGACTCC AGCCGAGCCC ACACCTGGCT ATGAGGTGGG CCAGCGGAAG CGCCTCATCT 420

CCTCGGTGGA GGACTTCACC GAGTTTGTGT GAGGCCGGGG CCCTCCCTCC TGCACTGGCC 480

TTGGACGGTA TTGCCTGTCA GTGAAATAAA TAAAGTCCTG ACCCCAGTGC ACAGACATAG 540

AGGCACAGAT TGC 553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6: (Compare Figure 9)

CTGGTGTGTG TGAGACGTGC GGGGCTGGGA AGTGTTGGCA GAGCCGCGAG TACCGTCCTC 60

ACTCCTTTTG TTCTTTTGAC GTAAGCTGGC GAGTGGCACT GCCTGAGTTC CGCTCAGTGC 120

CCGCCCTGAT GTGCGGACCC CGCTGCATTC TTGCTGTTAG GTGGTGGCGG TGTGCGCTGT 180

CGCTGGTGGG CACCGAGAGT CTTTGGGAGC TTTGGGGAGG TTGTGCCAAG CCTGAGCCTC 240

GACGTCCCCC TTCCCGGCTT TCTGTTGGCT CTTCTGAGGC CAGGGCATCT CTATGAGGGC 300

CTCCTGCTGG AGCCGTCTCT GTGGATCTCC TCTGCCATCC TGGCCCATGA GTGGGTGATG 360

CGCTGGCCAC CATCTGGTGA CAGTGGCCGG GCACCGCTGC CAAATGTGGG TCCCGCATCT 420

GCAAGCCCCT CCCTGGGTCC CCTAGGGTAT GGGGTGGTTC TGCCACTGCC CTCGCTCCCC 480

CACCTTGGGG TGCCTCTCCC CCTGCTCGTG GGGGAGA 517

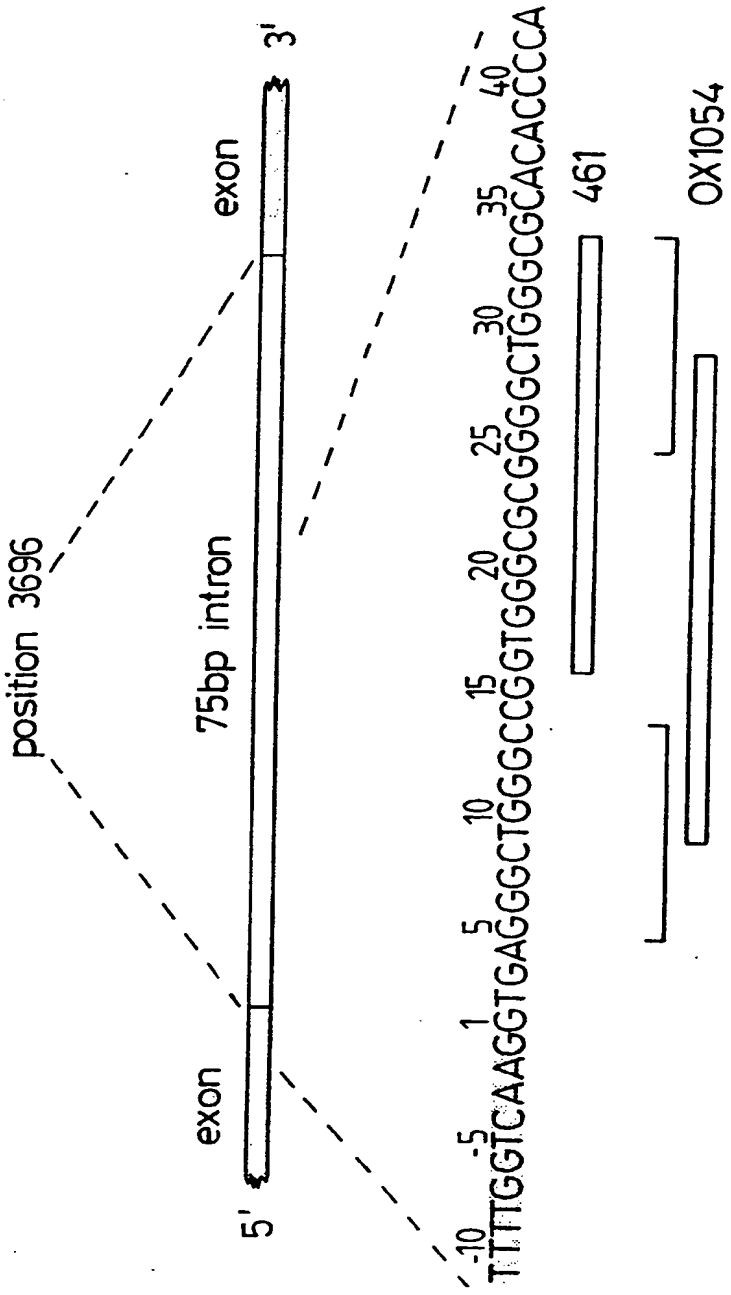

Fig. 11

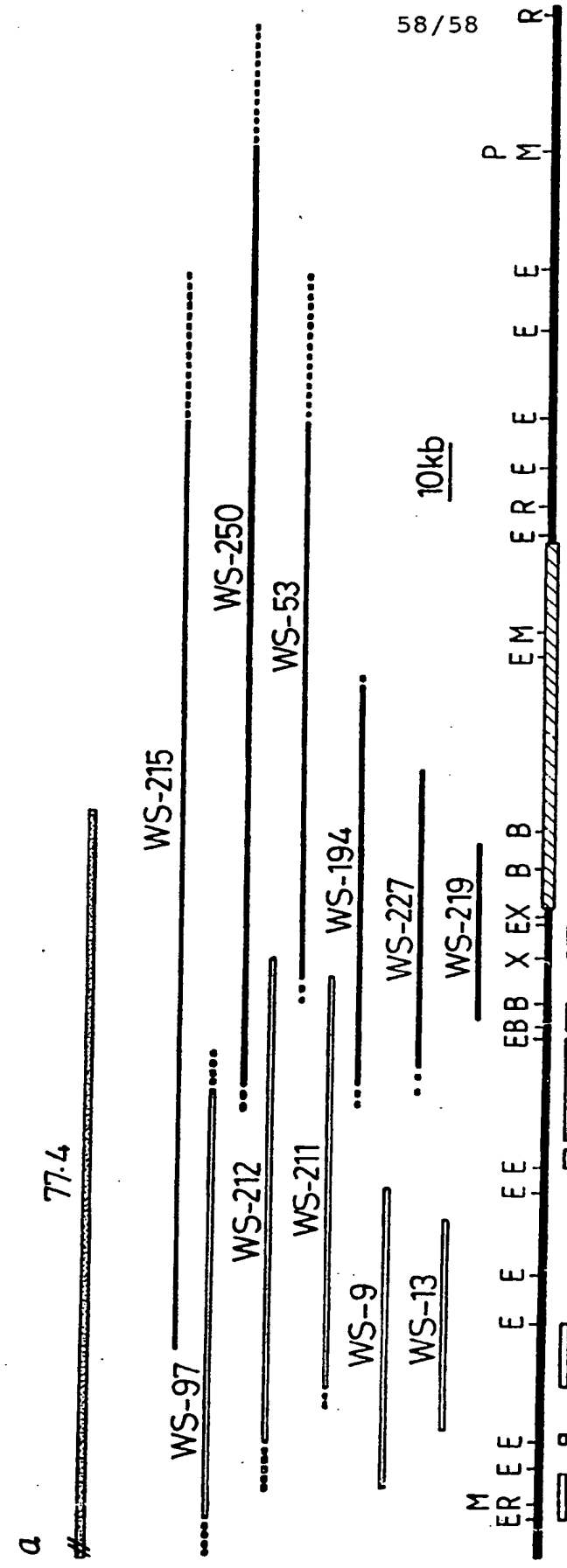
a
77.4
WS-215
WS-250
WS-97
WS-212
WS-53
WS-211
WS-194
WS-227
WS-219
WS-9
WS-13
10kb
M
ER E E E
SM9
CW9
E E E
CW15
H2
E E
CW18
CW20
EB B X EX B B
JH1
CW21
3'UTRP
JH6
JH8
JH11
CW36
CW10
E M
E R E E E E
SM6
JH17
P
M
R
TSC2
PKD1

Fig. 12

INTERNATIONAL SEARCH REPORT

Internat l Application No
PCT/GB 94/02822

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C12N5/10 A61K48/00 G01N33/68
C12Q1/68 C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A61K C12Q C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	J. AM. SOC. NEPHROL., vol. 4,no. 3, November 1993 page 814 G. GERMINO ET AL 'A novel approach to the identification of the PKD1 gene'	1-3,6-23
Y	see abstract 91p	24-30

Y	KIDNEY INTERNATIONAL, vol. 43,no. supp 39, 19 May 1993 pages s20-s25, G. GERMINO ET AL 'Positional cloning approach to the dominant polycystic kidney disease gene, PKD1' see the whole document	1-30

	-/--	

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 May 1995	19. 05. 95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internat ! Application No
PCT/GB 94/02822

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GENOMICS, vol. 13, 1992 pages 144-151, G. GERMINO ET AL 'The gene for autosomal dominant polycystic kidney disease.....' cited in the application see the whole document especially page 150, left column, last paragraph	1-30
Y	A. GRIFFITHS ET AL 'An introduction to genetic analysis' 1993 , W. FREEMAN AND COMPANY , NEW YORK see page 427 see page 453, left column, last paragraph - right column, paragraph 1 see page 453, right column, last paragraph - page 461	1-30
A	CURRENT OPINION IN GENETICS AND DEVELOPMENT, vol. 3, June 1993 pages 425-431, J. MULLEY ET AL 'Integrating maps of chromosome 16'	
X	EMBL DATABASE, Accession no. T05931, sequence reference HS9312, Sep. 2 1993; M. ADAMS et al 'Expressed sequence tags identify diversity of transcripts from human brain & NATURE GENETICS, vol. 4, 1993 pages 256-267,	1-3,6,8, 9
X	EMBL DATABASE, Accession no. T04943 sequence reference HS9431, August 30, 1993 M. ADAMS et al, 'Expressed sequence tags identify diversity of transcripts from human brain & NATURE GENETICS, vol. 4, 1993 pages 256-267,	1-3,6,8, 9
P,X	CELL, vol. 77, 17 June 1994 pages 881-894, C. WARD ET AL 'The polycystic kidney disease 1 gene encodes a 14kb transcript and lies within a duplicated region on chromosome 16' see the whole document	1-30

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB94/02822

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 24 partially and 25 are directed to methods of treatment of the human boby the search has been carried out and based on the alleged effect of the compound.

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.